

U -statistics of Ornstein-Uhlenbeck branching particle system

Radosław Adamczak* and Piotr Miłoś†

Faculty of Mathematics, Informatics and Mechanics, University of Warsaw
ul. Banacha 2, Warsaw, Poland

November 23, 2011

Abstract

We consider a branching particle system consisting of particles moving according to the Ornstein-Uhlenbeck process in \mathbb{R}^d and undergoing a binary, supercritical branching with a constant rate $\lambda > 0$. This system is known to fulfil a law of large numbers (under exponential scaling). Recently the question of the corresponding central limit theorem has been addressed. It turns out that the normalization and form of the limit in the CLT fall into three qualitatively different regimes, depending on the relation between the branching intensity and the parameters of the Ornstein-Uhlenbeck process. In the present paper we extend those results to U -statistics of the system proving a law of large numbers and a central limit theorem.

MSC: primary 60F05; 60J80 secondary 60G20

Keywords: Supercritical branching particle systems, U -statistics, Central limit theorem.

1 Introduction

We consider a single particle located at time $t = 0$ at $x \in \mathbb{R}^d$, moving according to the Ornstein-Uhlenbeck process and branching after exponential time independent of the spatial movement. The branching is supercritical and given by the generating function

$$F(s) := ps^2 + (1 - p), \quad p > \frac{1}{2}.$$

The offspring particles follow the same dynamics (independently of each other). We will refer to this system of particles as the OU branching process and denote it by $X = \{X_t\}_{t \geq 0}$.

Formally, we identify the system with the empirical process, i.e. X takes values in the space of Borel measures on \mathbb{R}^d and for each Borel set A , $X_t(A)$ is the (random) number of particles at time t in A .

*radamcz@mimuw.edu.pl. Research was partially supported by the MNiSW grant N N201 397437 and by the Foundation for Polish Science.

†pmilos@mimuw.edu.pl (corresponding author). Research was partially supported by the MNiSW grant N N201 397537.

It is well known (see e.g. [15]) that the system satisfies the law of large numbers, i.e. for all bounded continuous functions, conditionally on the set of non-extinction

$$|X_t|^{-1} \langle X_t, f \rangle \rightarrow \langle \varphi, f \rangle, \quad a.s., \quad (1)$$

where $|X_t|$ is the number of particles at time t , $\{X_t(1), X_t(2), \dots, X_t(|X_t|)\}$ are their positions, $\langle X_t, f \rangle := \sum_{i=1}^{|X_t|} f(X_t(i))$ and φ is the invariant measure of the Orstein-Uhlenbeck process.

In a recent article [1], we investigated second order behaviour of this system and proved central limit theorems corresponding to (1). It turns out that the behaviour of the system may fall into three qualitatively different categories, depending on the relation between the branching intensity and the parameters of the Orstein-Uhlenbeck process.

In the present article we extend these results on the LLN and CLT to the case of U -statistics of the system of arbitrary order $n \geq 1$, i.e. to random variables of the form

$$U_t^n(f) := \sum_{\substack{i_1, i_2, \dots, i_n=1 \\ i_k \neq i_j \text{ if } k \neq j}}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)) \quad (2)$$

(note that with this notation $\langle X_t, f \rangle$ corresponds to U -statistics of order $n = 1$). Our investigation parallels the classical and well-developed theory of U -statistics of independent random variables, however we would like to point out that in our context additional interest in this type of functionals of the process X stems from the fact that they capture 'average dependencies' between particles of the system. This will be seen from the form of the limit, which turns out to be more complicated than in the i.i.d. case.

The organization of the paper is as follows. After introducing the basic notation and preliminary facts in Section 2 we describe the main results of the paper in Section 3. Next (Section 4) we restate the results in the special case of $n = 1$ (as proven in [1]) to serve as a starting point for the general case. Finally, in Section 5 we provide complete proofs for arbitrary n . We conclude with some remarks concerning the so called non-degenerate case (Section 6).

2 Preliminaries

2.1 Notation

For a branching system $\{X_t\}_{t \geq 0}$, we denote by $|X_t|$ the number of particles at time t , and by $X_t(i)$ - the position of the i -th (in a certain ordering) particle at time t . We sometimes use \mathbb{E}_x or \mathbb{P}_x to denote the fact that we calculate the expectation for the system starting from a particle located at x . We use also \mathbb{E} and \mathbb{P} when this location is not relevant.

By \rightarrow^d we denote the convergence in law. We use \lesssim, \gtrsim to denote the situation when an equality or inequality holds with a constant $c > 0$, which is irrelevant for calculations. E.g. $f(x) \gtrsim g(x)$ means that there exists a constant $c > 0$ such that $f(x) = cg(x)$.

By $x \circ y = \sum_{i=1}^d x_i y_i$ we denote the standard scalar product of $x, y \in \mathbb{R}^d$, by $\|\cdot\|$ the corresponding Euclidean norm. By \otimes^n we denote the n -fold tensor product.

We use also $\langle f, \mu \rangle := \int_{\mathbb{R}^d} f(x) \mu(dx)$.

In our paper we will use the Feynman diagrams. A diagram γ labeled by $\{1, 2, \dots\}$ is a graph consisting of a set of edges E_γ not having common endpoints, and unpaired vertices A_γ . We will use $r(\gamma)$ to denote the rank of the diagram i.e. the number of edges. For properties and more information we refer to [17, Definition 1.35].

In the paper we will use the space

$$\mathcal{P} = \mathcal{P}(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \mapsto \mathbb{R} : f \text{ is continuous and } \exists_n \text{ such that } |f(x)|/\|x\|^n \rightarrow 0 \text{ as } \|x\| \rightarrow +\infty \right\}, \quad (3)$$

We endow this space with the following norm

$$\|f\|_{\mathcal{P}} = \|f\|_{\mathcal{P}(\mathbb{R}^d)} := \sup_{x \in \mathbb{R}^d} |n(x)f(x)|, \quad (4)$$

where $n(x) := \exp\left(-\sum_{i=1}^d |x_i|\right)$. We will also use

$$\mathcal{C}_c = \mathcal{C}_c(\mathbb{R}^d),$$

to denote the space of continuous compactly supported functions.

Given a function $f \in \mathcal{P}(\mathbb{R}^d)$ we will implicitly understand its derivatives (e.g. $\frac{\partial f}{\partial x_i}$) in the space of tempered distributions (see e.g. [21, p. 173]).

2.2 Basic facts on the Galton-Watson process

The number of particles $\{|X_t|\}_{t \geq 0}$ is the celebrated Galton-Watson process. We present basic properties of this process used in the paper. The main reference in this section is [2]. In our case the expected total number of particles grows exponentially at the rate

$$\lambda_p := (2p - 1)\lambda. \quad (5)$$

The process becomes extinct with probability (see [2, Theorem I.5.1])

$$p_e = \frac{1-p}{p}.$$

We will denote the extinction and non-extinction events by Ext and Ext^c respectively. The process $V_t := e^{-\lambda_p t} |X_t|$ is a positive martingale. Therefore it converges (see also [2, Theorem 1.6.1])

$$V_t \rightarrow V_\infty, \quad a.s. \quad \text{as } t \rightarrow +\infty. \quad (6)$$

We have the following simple fact (we refer to [1] for the proof).

Fact 2.1. *We have $\{V_\infty = 0\} = \text{Ext}$ and conditioned on non-extinction V_∞ has the exponential distribution with parameter $\frac{2p-1}{p}$. We have $\mathbb{E}(V_\infty) = 1$ and $\text{Var}(V_\infty) = \frac{1}{2p-1}$. $\mathbb{E}e^{-4\lambda_p t}|X_t|^4$ is uniformly bounded, i.e. there exists $C > 0$ such that for any $t \geq 0$ we have $\mathbb{E}e^{-4\lambda_p t}|X_t|^4 \leq C$. Moreover, all moments are finite, i.e. for any $n \in \mathbb{N}$ and $t \geq 0$ we have $\mathbb{E}|X_t|^n < +\infty$.*

We will denote the variable V_∞ conditioned on non-extinction by W .

2.3 Basic facts on the Ornstein-Uhlenbeck process

We recall that the Ornstein-Uhlenbeck process is a time homogenous Markov process with the infinitesimal operator

$$L := \frac{1}{2}\sigma^2\Delta - \mu x \circ \nabla. \quad (7)$$

The corresponding semigroup will be denoted by \mathcal{T} . The density of the invariant measure of the Ornstein-Uhlenbeck process is given by

$$\varphi(x) := \left(\frac{\mu}{\pi\sigma^2}\right)^{d/2} \exp\left(-\frac{\mu}{\sigma^2}\|x\|^2\right). \quad (8)$$

2.4 Basic facts concerning U -statistics

We will now briefly recall basic notation and facts concerning U -statistics. A U -statistic of degree n based on an \mathcal{X} -valued sample X_1, \dots, X_N and a function $f: \mathcal{X}^n \rightarrow \mathbb{R}$, is a random variable of the form

$$\sum_{\substack{i_1, i_2, \dots, i_n=1 \\ i_k \neq i_j \text{ if } k \neq j}}^N f(X_{i_1}, \dots, X_{i_n}).$$

The function f is usually referred to as the kernel of the U -statistic. Without loss of generality it can be assumed that f is symmetric i.e. invariant under permutation of its arguments. We refer the reader to [18, 7] for more information on U -statistics of sequences of independent random variables.

In our case we will consider U -statistics based on the sequence of positions of particles from the branching system as defined by (2). We will be interested in weak convergence of properly normalised U -statistics when $t \rightarrow \infty$. Similarly as in the classical theory, the asymptotic behaviour of U -statistics depends heavily on the so called order of degeneracy of the kernel f , which we will briefly recall in Section 5.1.2.

A function f is called completely degenerate or canonical (with respect to some measure of reference φ , which in our case will be the stationary measure of the Ornstein-Uhlenbeck process) if

$$\int_{\mathcal{X}} f(x_1, \dots, x_n) \varphi(dx_k) = 0,$$

for all $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n \in \mathcal{X}$. The complete degeneracy may be considered a centredness condition, in the classical theory of U -statistics canonical kernels are counterparts of centred random variables from the theory of sums of independent random variables. Their importance stems from the fact that each U -statistic can be decomposed into a sum of canonical U -statistics of different degrees, a fact known as the Hoeffding decomposition (see Section 5.1.2).

3 Main results

This section is devoted to the presentation of our results. The proofs are deferred to Section 5.

We start with the following law of large numbers (throughout the article when dealing with U -statistics of order n we will identify $\underbrace{\mathbb{R}^d \times \dots \times \mathbb{R}^d}_n$ with \mathbb{R}^{nd}).

Theorem 3.1. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f : \mathbb{R}^{nd} \mapsto \mathbb{R}$ is a bounded continuous function. Then, on the set of non-extinction Ext^c there is the convergence*

$$\lim_{n \rightarrow +\infty} |X_t|^{-n} U_t^n(f) = \langle f, \varphi^{\otimes n} \rangle \text{ a.s.} \quad (9)$$

Moreover, when $f \in \mathcal{P}(\mathbb{R}^{nd})$, then the above convergence holds in probability.

Having formulated the law of large numbers let us now pass to the corresponding CLTs. As already mentioned in the introduction, their form depends on the relation between λ_p and μ , more specifically we distinguish three cases $\lambda_p < 2\mu$, $\lambda_p = 2\mu$ and $\lambda_p > 2\mu$. We refer the reader to [1] for a detailed discussion of this phenomenon as well as its heuristic explanation and interpretation. Here we only stress that the situation for $\lambda_p > 2\mu$ differs substantially from the remaining two cases, as we obtain convergence in probability and the limit is not Gaussian even for $n = 1$ (intuitively, this is caused by large branching intensity which lets local correlations between particles prevail over the ergodic properties of the Orstein-Uhlenbeck process).

3.1 Slow branching case: $\lambda_p < 2\mu$

Let Z be a Gaussian stochastic measure on \mathbb{R}^{d+1} with intensity $\mu_1(dtdx) := (\delta_0(dt) + 2\lambda p e^{\lambda_p t} dt) \varphi(x) dx$ defined according to [17, Definition 7.17]. We denote the stochastic integral with respect to Z by I and the corresponding multiple stochastic integral by I_n [17, Section 7.2]. We assume that Z is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $f \in \mathcal{P}(\mathbb{R}^{nd})$ we define (we recall that \mathcal{T} is the semigroup of the Orstein-Uhlenbeck process)

$$H(f)(s_1, x_1, s_2, x_2, \dots, s_n, x_n) := (\otimes_{i=1}^n \mathcal{T}_{s_i}) f(x_1, x_2, \dots, x_n), \quad s_i \in \mathbb{R}_+, x_i \in \mathbb{R}^d. \quad (10)$$

It will be useful to treat this function as a function of n variables of type $z_i := (s_i, x_i) \in \mathbb{R}_+ \times \mathbb{R}^d$. For a Feynman diagram γ labeled by $\{1, 2, \dots, n\}$

$$L(f, \gamma) := I_{|A_\gamma|} \left(\left(\prod_{(j,k) \in E_\gamma} \int \mu_2(dz_{j,k}) \right) H(f)(u_1, u_2, \dots, u_n) \right), \quad \mu_2(dtdx) := 2\lambda p e^{\lambda_p t} \varphi(x) dt dx, \quad (11)$$

where $u_i = z_{j,k}$ if $(j,k) \in E_\gamma$ and $i = j$ or $i = k$ and $u_i = z_i$ if $i \in A_\gamma$. Less formally, for each pair (j,k) we integrate over diagonal of coordinates j and k with respect to μ_2 . The function obtained in this way

is integrated using the multiple stochastic integral $I_{|A_\gamma|}$. We define

$$L_1(f) := \sum_{\gamma} L(f, \gamma), \quad (12)$$

where the sum spans over all Feynman diagrams labeled by $\{1, 2, \dots, n\}$.

Fact 3.2. *For any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have $\mathbb{E}L_1(f)^2 < +\infty$. Moreover L_1 is a continuous function $L_1 : \text{Can} \mapsto L_2(\Omega, \mathcal{F}, \mathbb{P})$, where $\text{Can} = \{f \in \mathcal{P}(\mathbb{R}^{nd}) : f \text{ is a canonical kernel}\}$ and Can is endowed with the norm $\|\cdot\|_{\mathcal{P}}$.*

We are now ready to formulate our main result for processes with the small branching rate.

Theorem 3.3. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^{nd})$ is a canonical kernel and $\lambda_p < 2\mu$. Then conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{t\lambda_p} V_\infty}{\sqrt{|X_t|}}, \frac{U_t^n(f)}{|X_t|^{n/2}} \right) \rightarrow^d (W, G_1, L_1(f)), \quad (13)$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$ and $W, G_1, L_1(f)$ are independent random variables.

3.2 Critical branching case: $\lambda_p = 2\mu$

Consider the space $\mathcal{L} := L_2(\mathbb{R}^d, \Phi(dx))$, where

$$\Phi(x) := \sum_{l=1}^d \left| \frac{\partial \varphi(x)}{\partial x_l} \right|$$

and a centred Gaussian process $(G_f)_{f \in \mathcal{L}}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with the covariance structure given by

$$\text{Cov}(G_{f_1}, G_{f_2}) = \frac{\lambda_p \sigma^2}{\mu} \sum_{l=1}^d \left\langle f_1, \frac{\partial \varphi}{\partial x_l} \right\rangle \left\langle f_2, \frac{\partial \varphi}{\partial x_l} \right\rangle. \quad (14)$$

We will identify the process with a map $I : \mathcal{L} \rightarrow L_2(\Omega, \mathcal{F}, \mathbb{P})$, such that $I(f) = G_f$. One can easily check that I is a bounded linear operator.

To formulate the central limit theorem in this case we will need the following

Fact 3.4. *For every $n \geq 1$ there exists a unique bounded linear operator $L_2 : \text{Can} \mapsto L_2(\Omega, \mathcal{F}, \mathbb{P})$, where $\text{Can} = \{f \in \mathcal{P}(\mathbb{R}^{nd}) : f \text{ is a canonical kernel}\}$ endowed with the norm $\|\cdot\|_{\mathcal{P}}$, such that for every $f_1, \dots, f_n : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying $\langle f_i, \varphi \rangle = 0$ and $f_i \in \mathcal{P}(\mathbb{R}^d)$, $i = 1, \dots, n$,*

$$L_2(f_1 \otimes \dots \otimes f_n) = I(f_1) \cdots I(f_n). \quad (15)$$

The above lemma will be proved together with the following theorem, which describes the asymptotic behaviour of U -statistics in the critical case.

Theorem 3.5. Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^{nd})$ is a canonical kernel and $\lambda_p = 2\mu$. Then conditionally on the set of non-extinction Ext^c there is the convergence

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{t\lambda_p} V_\infty}{\sqrt{|X_t|}}, \frac{U_t^n(f)}{(t|X_t|)^{n/2}} \right) \rightarrow^d (W, G, L_2(f)),$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$ and $W, G, L_2(f)$ are independent random variables.

3.3 Fast branching case: $\lambda_p > 2\mu$

In order to describe the limit we introduce an \mathbb{R}^d -valued process

$$H_t := e^{(-\lambda_p + \mu)t} \sum_{i=1}^{|X_t|} X_t(i), \quad t \geq 0. \quad (16)$$

The following two facts have been proved in [1].

Fact 3.6. H is a martingale with respect to the filtration of the OU-branching system starting from $x \in \mathbb{R}^d$. Moreover for $\lambda_p > 2\mu$ we have $\sup_t \mathbb{E} H_t^2 < +\infty$, therefore there exists $H_\infty := \lim_{t \rightarrow +\infty} H_t$ (a.s. limit) and $H_\infty \in L_2$. When the OU branching system starts from 0 then martingales V_t and H_t are orthogonal.

It is worthwhile to note that the distribution of H_∞ depends on the starting conditions.

Fact 3.7. Let $\{X_t\}_{t \geq 0}$ and $\{\tilde{X}_t\}_{t \geq 0}$ be two OU branching processes, the first one starting from 0 and the second one from x . Let us denote the limit of corresponding martingales by $H_\infty, \tilde{H}_\infty$ respectively. Then

$$\tilde{H}_\infty =^d H_\infty + x V_\infty,$$

where V_∞ is the limit given by (6) for the system X .

H_∞ is \mathbb{R}^d -valued, we denote its coordinates by H_∞^i . Let $f \in \mathcal{P}(\mathbb{R}^{nd})$. We define

$$\tilde{L}_3(f) := \sum_{i_1, i_2, \dots, i_n=1}^d \left\langle \frac{\partial^n f}{\partial x_{1,i_1} \partial x_{2,i_2} \dots \partial x_{n,i_n}}, \varphi^{\otimes n} \right\rangle H_\infty^{i_1} H_\infty^{i_2} \dots H_\infty^{i_n}.$$

where we adopted convention that $x_{j,l}$ is the l -th coordinate of the j -th variable. By $L_3(f)$ we will denote $\tilde{L}_3(f)$ conditioned on Exp^c .

Theorem 3.8. Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^{nd})$ is a canonical kernel and $\lambda_p > 2\mu$. Then conditionally on the set of non-extinction Ext^c there is the convergence

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{t\lambda_p} V_\infty}{\sqrt{|X_t|}}, e^{-n(\lambda_p - \mu)t} U_t^n(f) \right) \rightarrow^d (W, G_1, L_3(f)), \quad (17)$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$ and $(W, L_3(f)), G_1$ are independent. Moreover

$$\left(e^{-\lambda_p t} |X_t|, e^{-n(\lambda_p - \mu)t} U_t^n(f) \right) \rightarrow^d \left(V_\infty, \tilde{L}_3(f) \right), \quad \text{in probability.}$$

3.4 Remarks on the CLT for U -statistics of i.i.d. random variables

For comparison purposes we will now briefly recall known results on the central limit theorem for U -statistics of independent random variables. U -statistics were introduced in the 1940's in the context of unbiased estimation by Halmos [14] and Hoeffding who obtained the central limit theorem for non-degenerate (degenerate of order 0) kernels [16]. The full description of the central limit theorem was obtained in [20, 11] (see also the article [13] where the CLT is proven for a related class of V -statistics). Similarly as in our case, the asymptotic behaviour of U -statistics based on a function $f: \mathcal{X}^n \rightarrow \mathbb{R}$ and an i.i.d. \mathcal{X} -valued sequence X_1, X_2, \dots is governed by the order of degeneracy of the function f (see Section 5.1.2) with respect to the law of X_1 (call it P). The case of general f can be reduced to the canonical one, for which one has the weak convergence

$$N^{-n/2} \sum_{\substack{i_1, i_2, \dots, i_n=1 \\ i_k \neq i_j \text{ if } k \neq j}}^N f(X_{i_1}, \dots, X_{i_n}) \rightarrow J_n(f),$$

where J_n is the n -fold stochastic integral with respect to the so-called isonormal process on \mathcal{X} , i.e. the stochastic Gaussian measure with intensity P .

Let us note that in the i.i.d. case the limiting distribution is simpler than for U -statistics of the OU branching processes. For small branching rate however, the behaviour of U -statistics in our case still resembles the classical one as it is a sum of multiple stochastic integrals of different orders. In the remaining two cases the behaviour differs substantially. This can be seen as a result of the lack of independence. Although asymptotically the particles' positions become less and less dependent, in short time scale offspring of the same particle stay close one to another.

Let us finally mention some results for U -statistics in dependent situations, which have been obtained in the last years. In [6] the authors analysed the behaviour of U -statistics of stationary absolutely regular sequences and obtained the CLT in the non-degenerate case (with Gaussian limit). In [5] the authors considered α and φ mixing sequences and obtained a general CLT for canonical kernels. Interesting results for long-range dependent sequences have been also obtained in [8]. A more recent interesting work is [19], where the authors consider U -statistics of interacting particle systems.

4 The case of $n = 1$

In the special case of $n = 1$ the results presented in the previous section were proven in [1]. Although this case obviously follows immediately from the results for general n it is actually a starting point in the proof of the general result (similarly as in the case of U -statistics of i.i.d. random variables). Therefore, for the reader's convenience, we will now restate this case in a simpler language of [1], not involving multiple stochastic integrals.

We will start with the law of large numbers

Theorem 4.1. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^d)$. Then*

$$\lim_{t \rightarrow +\infty} e^{-\lambda_p t} \langle X_t, f \rangle = \langle f, \varphi \rangle V_\infty \quad \text{in probability,}$$

or equivalently on the set of non-extinction, Ext^c , we have

$$\lim_{t \rightarrow +\infty} |X_t|^{-1} \langle X_t, f \rangle = \langle f, \varphi \rangle \quad \text{in probability.} \quad (18)$$

Moreover, if f is bounded then the almost sure convergence holds.

4.1 Small branching rate: $\lambda_p < 2\mu$

We denote $\tilde{f}(x) := f(x) - \langle f, \varphi \rangle$ and

$$\sigma_f^2 := \left\langle \varphi, \tilde{f}^2 \right\rangle + 2\lambda_p \int_0^{+\infty} \left\langle \varphi, \left(e^{(\lambda_p/2)s} \mathcal{T}_s \tilde{f} \right)^2 \right\rangle ds. \quad (19)$$

Let us also recall (6) and that W is V_∞ conditioned on Ext^c . In this case, the behaviour of X is given by the following

Theorem 4.2. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $\lambda_p < 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^d)$. Then $\sigma_f^2 < +\infty$ and conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{t\lambda_p} V_\infty}{\sqrt{|X_t|}}, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{\sqrt{|X_t|}} \right) \rightarrow^d (W, G_1, G_2),$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$, $G_2 \sim \mathcal{N}(0, \sigma_f^2)$ and W, G_1, G_2 are independent random variables.

4.2 Critical branching rate: $\lambda_p = 2\mu$

We denote

$$\sigma_f^2 := \frac{\lambda_p \sigma^2}{\mu} \sum_{i=1}^d \left\langle f, \frac{\partial \varphi}{\partial x_i} \right\rangle^2. \quad (20)$$

Note that the same symbol σ_f^2 has already been used to denote the asymptotic variance in the small branching case. However, since these cases will always be treated separately, this should not lead to ambiguity.

Theorem 4.3. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $\lambda_p = 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^d)$. Then $\sigma_f^2 < +\infty$ and conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{t\lambda_p} V_\infty}{\sqrt{|X_t|}}, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{t^{1/2} \sqrt{|X_t|}} \right) \rightarrow^d (W, G_1, G_2),$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$, $G_2 \sim \mathcal{N}(0, \sigma_f^2)$ and W, G_1, G_2 are independent random variables.

4.3 Fast branching rate: $\lambda_p > 2\mu$

In the following theorem we use the notation introduced in Section 3.3.

Theorem 4.4. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $\lambda_p > 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^d)$. Then conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{t\lambda_p} V_\infty}{\sqrt{|X_t|}}, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{\exp((\lambda - \mu)t)} \right) \rightarrow^d (W, G, \langle \text{grad } f, \varphi \rangle \circ J), \quad (21)$$

where $G \sim \mathcal{N}(0, 1/(2p-1))$, $(W, J), G$ are independent and J is H_∞ conditioned on Ext^c . Moreover

$$\left(e^{-\lambda_p t} |X_t|, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{\exp((\lambda - \mu)t)} \right) \rightarrow (V_\infty, \langle \text{grad } f, \varphi \rangle \circ H_\infty), \quad \text{in probability.}$$

5 Proofs

We will now pass to the proofs of the results announced in Section 3. Their general structure is similar as in the case of U -statistics of independent random variables, i.e. all the theorems will be proved first for linear combinations of tensor products and then via suitable approximations extended to proper function spaces.

In the next section we will recall some additional (standard) facts concerning the Orstein-Uhlenbeck process and U -statistics. Next, in Section 5.2 we will develop general tools needed for the approximation, which will be the most technical part of the proof. Section 5.3 will be devoted to rather short proofs of the main results.

From now on we will often work conditionally on the set of non-extinction Ext^c , which will not be explicitly mentioned in the proofs (however should be clear from the context).

5.1 Auxiliary facts and notation

5.1.1 The Orstein-Uhlenbeck process

The semigroup of the Ornstein-Uhlenbeck process can be represented by

$$\mathcal{T}_t f(x) = (g_t * f)(x_t), \quad x_t := e^{-\mu t} x, \quad (22)$$

where

$$g_t(x) = \left(\frac{\mu}{\pi \sigma_t^2} \right)^{d/2} \exp \left\{ -\frac{\mu}{\sigma_t^2} \|x\|^2 \right\}, \quad \sigma_t^2 := \sigma^2 (1 - e^{-2\mu t}).$$

Let us recall (8). We denote $ou(t) := \sqrt{1 - e^{-2\mu t}}$ and let $G \sim \varphi$. The semigroup \mathcal{T} has the following useful representations

$$\mathcal{T}_t f(x) = \int_{\mathbb{R}^d} f(x_t - y) g_t(y) dy = \int_{\mathbb{R}^d} f(x e^{-\mu t} + ou(t)y) \varphi(y) dy = \mathbb{E} f(x e^{-\mu t} + ou(t)G). \quad (23)$$

We also denote

$$\mathcal{T}_s^\lambda := e^{\lambda s} \mathcal{T}_s.$$

5.1.2 U - and V -statistics

We will now briefly recall one of the standard tools of the theory of U -statistics, which we will use in the sequel, namely the Hoeffding decomposition.

Let us introduce for $I \subseteq \{1, \dots, n\}$ the Hoeffding projection of f corresponding to I as the function $\Pi_I f: \mathbb{R}^{|I|d} \rightarrow \mathbb{R}$, given by the formula

$$\Pi_I f((x_i)_{i \in I}) = \int_{\mathbb{R}^{nd}} \left(\prod_{i \notin I} \varphi(dy_i) \prod_{i \in I} (\delta_{x_i} - \varphi)(dy_i) \right) f(y_1, \dots, y_n). \quad (24)$$

Once can easily see that for $|I| \geq 1$, $\Pi_I f$ is a canonical kernel. Moreover $\Pi_\emptyset f = \int_{\mathbb{R}^{nd}} f(x_1, \dots, x_n) \prod_{i=1}^n \varphi(dx_i)$.

Note that if f is symmetric (i.e. invariant with respect to permutations of arguments), $\Pi_I f$ depends only on the cardinality of f . In this case we speak about the k -th Hoeffding projection ($k = 0, \dots, n$), given by

$$\Pi_k f(x_1, \dots, x_k) = \left\langle (\delta_{x_1} - \varphi) \otimes \dots \otimes (\delta_{x_k} - \varphi) \otimes \varphi^{\otimes(n-k)}, f \right\rangle.$$

A symmetric kernel in n variables is called *degenerate of order $k - 1$* ($1 \leq k \leq n$) iff $k = \min\{i > 0: \Pi_i f \neq 0\}$. The order of degeneracy is responsible for the normalisation and the form of the limit in the central limit theorem for U -statistics, e.g. if the kernel is non-degenerate, i.e. $\Pi_1 f \neq 0$, then the corresponding U -statistic of an i.i.d. sequence behaves like a sum of independent random variables and converges to a Gaussian limit. The same phenomenon will be present also in our situation (see Section 6).

In the particular case $k = n$ the definition of the Hoeffding projection reads as

$$\Pi_n(f) := \langle (\otimes_{i=1}^n (\delta_{x_i} - \varphi)), f \rangle.$$

One easily checks that

$$f(x_1, \dots, x_n) = \sum_{I \subseteq \{1, \dots, n\}} \Pi_I f((x_i)_{i \in I}),$$

which gives us the aforementioned Hoeffding decomposition of U -statistics

$$U_t^n(f) = \sum_{I \subseteq \{1, \dots, n\}} \frac{(|X_t| - |I|)!}{(|X_t| - n)!} U_t^{|I|}(\Pi_I f),$$

which in the case of symmetric kernels simplifies to

$$U_t^n(f) = \sum_{k=0}^n \binom{n}{k} \frac{(|X_t| - k)!}{(|X_t| - n)!} U_t^k(\Pi_k f),$$

where we use the convention $U_t^0(a) = a$ for any constant a .

For technical reasons we will also consider the notion of a V -statistic which is closely related to U -statistics, and is defined as

$$V_t^n(f) := \sum_{i_1, i_2, \dots, i_n=1}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)). \quad (25)$$

The corresponding Hoeffding decomposition is

$$V_t^n(f) = \sum_{I \subseteq \{1, \dots, n\}} |X_t|^{n-|I|} V_t^{|I|}(\Pi_I f), \quad (26)$$

where again we set $V_t^0(a) = a$ for any constant a .

In the proof of our results we will use a standard observation that a U -statistic can be written as a sum of V -statistics. More precisely, let \mathcal{J} be the collection of partitions of $\{1, \dots, n\}$ i.e. of all sets $J = \{J_1, \dots, J_k\}$, where J_i 's are nonempty, pairwise disjoint and $\bigcup_i J_i = \{1, \dots, n\}$. For J as above let f_J be a function of $|J|$ variables $x_1, \dots, x_{|J|}$, obtained by substituting x_i for all the arguments of f corresponding to the set J_i , e.g. for $n = 3$ and $J = \{\{1, 2\}, \{3\}\}$, $f_J(x_1, x_2) = f(x_1, x_1, x_2)$. An easy application of the inclusion-exclusion formula yields that

$$U_t^n(f) = \sum_{J \in \mathcal{J}} a_J V_t^{|J|}(f_J), \quad (27)$$

where a_J are some integers depending only on the partition J . Moreover one can easily check that if $J = \{\{1\}, \dots, \{n\}\}$, then $a_J = 1$, whereas if J consists of sets with at most two elements then $a_J = (-1)^k$ where k is the number of two-element sets in J . Let us also note that partitions consisting only of one- and two-element sets can be in a natural way identified with the Feynman diagrams (defined in Section 2.1).

5.2 Approximation

5.2.1 Approximation of functions

First we will show that any function in $\mathcal{P}(\mathbb{R}^{nd})$ can be approximated by tensor functions. For a subset A of a linear space by $\text{span}(A)$ we denote the set of finite linear combinations of elements of A .

Lemma 5.1. *Let $A := \{\otimes_{i=1}^n f_i : f_i \text{ bounded continuous}\}$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ be a canonical kernel then there exists a sequence $\{f_k\} \subset \text{span}(A)$ such that each f_k is canonical and*

$$\|f_k - f\|_{\mathcal{P}} \rightarrow 0, \quad \text{as } k \rightarrow +\infty.$$

Proof. First we prove that $\text{span}(A)$ is dense in $\mathcal{P}(\mathbb{R}^{nd})$. Let us notice that given a function $f \in \mathcal{P}(\mathbb{R}^{nd})$ it suffices to approximate it on some box $[-M, M]^d$, $M > 0$. The box is a compact set and an approximation exists due to the Stone-Weierstrass theorem.

Now, let $f \in \mathcal{P}(\mathbb{R}^{nd})$. We may find a sequence $\{h_k\} \subset \text{span}(A)$ such that $h_k \rightarrow f$ in \mathcal{P} . Let us recall the Hoeffding projection (24) and denote $I = \{1, 2, \dots, n\}$. Now direct calculation (using the exponential integrability of Gaussian variables) reveals that the sequence $f_k := \Pi_I h_k$ fulfils the conditions of the lemma. \square

Let $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $I \subset \{1, 2, \dots, n\}$ with $|I| = k$. We define

$$\hat{f}_I(x_1, x_2, \dots, x_n) := \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i, \quad (28)$$

where $z_i = y_i$ if $i \in I$, $z_i = x_i$ otherwise and g_1 is given by (22). We have

Lemma 5.2. *Let $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $l \in \mathbb{N}$. Then for any $I \subset \{1, 2, \dots, n\}$ the function \hat{f}_I is smooth with respect to coordinates in I . For any multi-index $\Lambda = (i_1, \dots, i_l) \subset \{1, \dots, nd\}^l$ such that $\{\lceil i_j/d \rceil : j = 1, \dots, l\} \subset I$ we have*

$$\left\| \frac{\partial^{|\Lambda|} \hat{f}_I}{\partial x_\Lambda} \right\|_{\mathcal{P}} \leq C \|f\|_{\mathcal{P}},$$

where $C > 0$ depends only on σ, μ, d, l, n . Moreover, when f is canonical, so is \hat{f}_I .

Proof. Let us fix some I and Λ . Using (28) we get

$$K := \frac{\partial^{|\Lambda|} \hat{f}_I}{\partial x_\Lambda} = \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \frac{\partial^{|\Lambda|}}{\partial x_\Lambda} \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i.$$

Therefore by the properties of Gaussian density g_1 and easy calculations we arrive at

$$|K| \leq \|f\|_{\mathcal{P}} \int_{\mathbb{R}^{kd}} n(z_1, z_2, \dots, z_n)^{-1} \left| \frac{\partial^{|\Lambda|}}{\partial x_\Lambda} \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) \right| dy_i \leq C_\Lambda n(x_1, x_2, \dots, x_n)^{-1},$$

for some constant C_Λ .

To conclude it is enough to take the maximum over all admissible pairs I, Λ .

Let us now assume that f is canonical. We would like to check that for any $j \in \{1, 2, \dots, n\}$ we have

$$\int_{\mathbb{R}^{nd}} \hat{f}(x_1, x_2, \dots, x_n) \varphi(x_j) dx_j = 0.$$

There are two cases, the first when $j \notin I$. Then we have

$$\begin{aligned} \int_{\mathbb{R}^d} \hat{f}(x_1, x_2, \dots, x_n) \varphi(x_j) dx_j &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i \varphi(z_j) dz_j \\ &= \int_{\mathbb{R}^{kd}} \left(\int_{\mathbb{R}^d} f(z_1, z_2, \dots, z_n) \varphi(z_j) dz_j \right) \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i = 0. \end{aligned}$$

The second case is when $j \in I$. Then

$$\begin{aligned} \int_{\mathbb{R}^d} \hat{f}(x_1, x_2, \dots, x_n) \varphi(x_j) dx_j &= \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I \setminus \{j\}} g_1(x_i e^{-\mu} - y_i) dy_i \left(\int_{\mathbb{R}^d} g_1(x_j e^{-\mu} - y_j) \varphi(x_j) dx_j \right) dy_j \\ &= \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I \setminus \{j\}} g_1(x_i e^{-\mu} - y_i) dy_i \varphi(y_j) dy_j, \end{aligned}$$

where the second equality holds by the fact that φ is the invariant measure of the Ornstein-Uhlenbeck process. Now the proof reduces to the first case. \square

We will also need the following simple identity. We consider $\{x_i\}_{i=1,2,\dots,n}, \{\tilde{x}_i\}_{i=1,2,\dots,n}$. By induction one easily checks (we slightly abuse the notation here, e.g. $\frac{\partial}{\partial y_i}$ denotes the derivative in direction $\tilde{x}_i - x_i$ and \int_a^b the integral over the segment $[a, b] \subset \mathbb{R}^d$)

Lemma 5.3. *Let f be a smooth function, then*

$$\sum_{(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \in \{0,1\}^n} (-1)^{\sum_{i=1}^n \epsilon_i} f(\tilde{x}_1 + \epsilon_1(x_1 - \tilde{x}_1), \tilde{x}_2 + \epsilon_2(x_2 - \tilde{x}_2), \dots, \tilde{x}_n + \epsilon_n(x_n - \tilde{x}_n)) = \int_{x_1}^{\tilde{x}_1} \int_{x_2}^{\tilde{x}_2} \dots \int_{x_n}^{\tilde{x}_n} \frac{\partial^n}{\partial y_1 \partial y_2 \dots \partial y_n} f(y_1, y_2, \dots, y_n) dy_n dy_{n-1} \dots dy_1.$$

5.2.2 Approximation of U -statistics

Bookkeeping of trees So far we have shown that one can approximate functions in \mathcal{P} by linear combinations of tensor products. Our next goal is to show that two functions which are close in \mathcal{P} generate U -statistics which (after proper normalization, specific for each regime) are close in distribution. To this end we will use the “bookkeeping of trees” technique (see e.g. [4] or [10, Section 2]), which via some combinatorics and introduction of auxiliary branching processes will allow us to pass from equations on the Laplace transform in the case of $n = 1$ to estimates of moments of V -statistics and consequently U -statistics.

We recall (25). Let $f_1, f_2, \dots, f_n \in \mathcal{C}_c(\mathbb{R}^d)$ and $f_i \geq 0$. We would like to calculate

$$\mathbb{E}_x V_t^n(\otimes_{i=1}^n f_i) = \mathbb{E}_x \left(\prod_{i=1}^n \langle X_t, f_i \rangle \right). \quad (29)$$

Let $\Lambda \subset \{1, 2, \dots, n\}$, slightly abusing notation we denote

$$w_\Lambda(x, t, \alpha) := \mathbb{E}_x \left(\exp \left\{ - \sum_{i \in \Lambda} \alpha_i \langle X_t, f_i \rangle \right\} \right), \quad w_\Lambda(x, t) := \left. \frac{\partial^{|\Lambda|}}{\partial \alpha_\Lambda} w(x, t, \alpha) \right|_{\alpha=0}.$$

Note that this differentiation is valid by Fact 2.1 and properties of the Laplace transform (e.g. [12, Chapter XIII.2]). By the calculations from Section 4.2. in [1] we know that

$$w_\Lambda(x, t, \alpha) = \mathcal{T}_t e^{-\sum_{i \in \Lambda} \alpha_i f_i}(x) + \lambda \int_0^t \mathcal{T}_{t-s} [p w_\Lambda^2(\cdot, s, \alpha) - w_\Lambda(\cdot, s, \alpha) + (1-p)] ds.$$

It is easy to check that

$$v_\Lambda(x, t) := \mathbb{E}_x \left(\prod_{i \in \Lambda} \langle X_t, f_i \rangle \right) = (-1)^{|\Lambda|} w_\Lambda(x, t). \quad (30)$$

Assume that $|\Lambda| > 0$. We denote by $P_1(\Lambda)$ all pairs (Λ_1, Λ_2) such that $\Lambda_1 \cup \Lambda_2 = \Lambda$ and $\Lambda_1 \cap \Lambda_2 = \emptyset$, and by $P_2(\Lambda) \subset P_1(\Lambda)$ pairs with an additional restriction that $\Lambda_1 \neq \emptyset$ and $\Lambda_2 \neq \emptyset$. We easily check that

$$\begin{aligned} \frac{\partial^{|\Lambda|}}{\partial \alpha_\Lambda} w_\Lambda(x, t, \alpha) &= (-1)^{|\Lambda|} \mathcal{T}_t \left(\prod_{i \in \Lambda} f_i e^{-\sum_{i \in \Lambda} \alpha_i f_i} \right) (x) \\ &+ \lambda \int_0^t \mathcal{T}_{t-s} \left[p \sum_{(\Lambda_1, \Lambda_2) \in P_1(\Lambda)} \frac{\partial^{|\Lambda_1|}}{\partial \alpha_{\Lambda_1}} w_{\Lambda_1}(\cdot, s, \alpha) \frac{\partial^{|\Lambda_2|}}{\partial \alpha_{\Lambda_2}} w_{\Lambda_2}(\cdot, s, \alpha) - \frac{\partial^{|\Lambda|}}{\partial \alpha_\Lambda} w_\Lambda(\cdot, s, \alpha) \right] (x) ds. \end{aligned}$$

We evaluate it at $\alpha = 0$, (let us notice that $\frac{\partial^{|\Lambda_1|}}{\partial \alpha_{\Lambda_1}} w_\Lambda(x, s, 0) = \frac{\partial^{|\Lambda_1|}}{\partial \alpha_{\Lambda_1}} w_{\Lambda_1}(x, s, 0) = w_{\Lambda_1}(x, s)$), multiply both sides by $(-1)^{|\Lambda|}$ and use the definition of $v_\Lambda(x, t)$ to get

$$\begin{aligned} v_\Lambda(x, t) &= \mathcal{T}_t \left(\prod_{i \in \Lambda} f_i \right) (x) + \lambda \int_0^t \mathcal{T}_{t-s} \left[p \sum_{(\Lambda_1, \Lambda_2) \in P_1(\Lambda)} v_{\Lambda_1}(\cdot, s) v_{\Lambda_2}(\cdot, s) - v_\Lambda(\cdot, s) \right] (x) ds \\ &= \mathcal{T}_t \left(\prod_{i \in \Lambda} f_i \right) (x) + \lambda \int_0^t \mathcal{T}_{t-s} \left[p \sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} v_{\Lambda_1}(\cdot, s) v_{\Lambda_2}(\cdot, s) + (2p - 1) v_\Lambda(\cdot, s) \right] (x) ds. \end{aligned}$$

This can be easily transformed to (recall that $\mathcal{T}_s^{\lambda p} f(x) = e^{\lambda p s} \mathcal{T}_s f(x)$)

$$v_\Lambda(x, t) = \mathcal{T}_t^{\lambda p} \left(\prod_{i \in \Lambda} f_i \right) (x) + p\lambda \int_0^t \mathcal{T}_{t-s}^{\lambda p} \left[\sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} v_{\Lambda_1}(\cdot, s) v_{\Lambda_2}(\cdot, s) \right] (x) ds. \quad (31)$$

The last formula is much easier to handle if written in terms of auxiliary branching processes. Firstly, we introduce the following notation. For $n \in \mathbb{N} \setminus \{0\}$ we denote by \mathbb{T}_n the set of rooted trees described below. The root has a single offspring. All inner vertices (we exclude the root and the leaves) have exactly two offspring. For $\tau \in \mathbb{T}_n$, by $l(\tau)$ we denote the set of its leaves. Each leaf $l \in l(\tau)$ is assigned a label, denoted by $\text{lab}(l)$, which is a non-empty subset of $\{1, 2, \dots, n\}$, the labels fulfil two conditions:

$$\bigcup_{l \in l(\tau)} \text{lab}(l) = \{1, 2, \dots, n\}, \quad \forall_{l_1, l_2 \in l(\tau)} (l_1 \neq l_2 \implies \text{lab}(l_1) \cap \text{lab}(l_2) = \emptyset).$$

In other words, the labels form a partition of $\{1, 2, \dots, n\}$. For a given $\tau \in \mathbb{T}_n$ let $i(\tau)$ denote the set of inner vertices (we exclude the root and the leaves), clearly $|i(\tau)| = |l(\tau)| - 1$ (as usual $|\cdot|$ denotes the cardinality). Let us identify the vertices of τ with $\{0, 1, 2, \dots, |\tau| - 1\}$ in such way, that for any vertex i its parent, denoted by $p(i)$, is smaller. Obviously, this implies that 0 is the root and that the inner vertices have numbers in set $\{1, 2, \dots, |\tau| - 1\}$. We denote also

$$s(\tau) := \{l \in l(\tau) : |\text{lab}(l)| = 1\}, \quad m(\tau) := \{l \in l(\tau) : |\text{lab}(l)| > 1\}, \quad (32)$$

leaves with single and multiple labels respectively.

Let $\tau \in \mathbb{T}_n$, we consider an Ornstein-Uhlenbeck branching walk on τ as follows (in this part we ignore the labels). Let us fix $t \in \mathbb{R}_+$ and $\{t_i\}_{i \in i(\tau)}$. The initial particle is placed at time 0 at location x , it evolves up to the time $t - t_1$ and splits into two offspring, the first one is associated with the left branch of vertex 1 in tree τ and the second one with the right branch. Further each of them evolves until time $t - t_i$, where i is the first vertex in the corresponding subtree, when it splits and so on. At time t the particles are stopped and their positions are denoted by $\{Y_i\}_{i \in l(\tau)}$ (the number of particles at the end is equal to the number of leaves). The construction makes sense provided that $t_i \leq t$ and $t_i \leq t_{p(i)}$ for all $i \in i(\tau)$. We define

$$OU \left(\bigotimes_{i=1}^n f_i, \tau, t, \{t_i\}_{i \in i(\tau)}, x \right) := \mathbb{E} \left(\prod_{a=1}^n f_a(Y_{j(a)}) \right), \quad (33)$$

where $j(a) = l \in l(\tau)$ is the unique leaf such that $a \in \text{lab}(l)$. We also define

$$S(\tau, t, x) := (p\lambda)^{|i(\tau)|} e^{\lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) OU \left(\otimes_{i=1}^n f_i, \tau, t, \{t_i\}_{i \in i(\tau)}, x \right). \quad (34)$$

Fact 5.4. *Let $\Lambda_n = \{1, 2, \dots, n\}$. The following identity holds*

$$v_{\Lambda_n}(x, t) = \sum_{\tau \in \mathbb{T}_n} S(\tau, t, x). \quad (35)$$

Proof. The claim is a consequence of the identity $v_{\Lambda}(x, t) = \sum_{\tau \in \mathbb{T}(\Lambda)} S(\tau, t, x)$, where $\Lambda \subset \{1, 2, \dots, n\}$ and $\mathbb{T}(\Lambda)$ is the set of trees, as $\mathbb{T}_{|\Lambda|}$, with the exception that the labels are in the set Λ . This identity in turn will follow by induction with respect to the cardinality of Λ . For $\Lambda = \{i\}$ equation (31) reads as $v_{\Lambda}(x, t) = e^{\lambda_p t} \mathcal{T}_t f_i(x)$. The space $\mathbb{T}(\Lambda)$ contains only one tree, denoted by τ_s , consisting of the root and a single leaf labelled by i . We have $i(\tau) = \emptyset$ and obviously $S(\tau_s, t, x) = e^{\lambda_p t} \mathcal{T}_t f_i(x)$. Let now $|\Lambda| = k > 1$. Similarly as before the first term of (31) corresponds to τ_s . By induction the second term can be written as

$$e^{\lambda_p t} p\lambda \int_0^t e^{-\lambda_p t_1} \mathcal{T}_{t-t_1} \left[\sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} \sum_{\tau_1 \in \mathbb{T}(\Lambda_1)} \sum_{\tau_2 \in \mathbb{T}(\Lambda_2)} S(\tau_1, t_1, \cdot) S(\tau_2, t_1, \cdot) \right] (x) dt_1.$$

We have

$$\begin{aligned} & \sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} \sum_{\tau_1 \in \mathbb{T}(\Lambda_1)} \sum_{\tau_2 \in \mathbb{T}(\Lambda_2)} e^{\lambda_p t} p\lambda \int_0^t \int_{\mathbb{R}^d} e^{-\lambda_p t_1} p(t - t_1, x, y) \left[\right. \\ & (p\lambda)^{|i(\tau_1)|} e^{\lambda_p t_1} \left(\prod_{i \in i(\tau_1)} \int_0^{t_1} dt_i \right) \left(\prod_{i \in i(\tau_1)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) OU \left(\otimes_{i=1}^n f_i, \tau_1, t_1, \{t_i\}_{i \in i(\tau_1)}, y \right) \\ & \left. (p\lambda)^{|i(\tau_2)|} e^{\lambda_p t_1} \left(\prod_{i \in i(\tau_2)} \int_0^{t_1} dt_i \right) \left(\prod_{i \in i(\tau_2)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) OU \left(\otimes_{i=1}^n f_i, \tau_2, t_1, \{t_i\}_{i \in i(\tau_2)}, y \right) \right] dy dt_1, \end{aligned}$$

where p is the transition density of the Ornstein-Uhlenbeck process. Now we create a new tree τ by setting τ_1 and τ_2 to be descendants of the vertex born at time $t - t_1$. We keep labels and split times unchanged and assume that the first particles of OU processes on τ_1 and τ_2 are put at x . Thus by the Markov property of the Ornstein-Uhlenbeck process we can identify the branching random walk on τ_1 and τ_2 with the branching random walk on τ . It is also easy to check that the described correspondence is a bijection from set of pairs (τ_1, τ_2) (as in the sum above) to $\mathbb{T}_n \setminus \{\tau_s\}$ and therefore the expression above is equal to the sum $\sum_{\tau \in \mathbb{T}_n \setminus \{\tau_s\}} S(\tau, t, x)$. \square

The calculations will be more tractable when we derive an explicit formula for $\{Y_i\}_{i \in l(\tau)}$. Let us recall the notation introduced in (23) and consider a family of independent random variables $\{G_i\}_{i \in \tau}$, such that $G_i \sim \varphi$ for $i \neq 0$ and $G_0 \sim \delta_x$. Recall also that $ou(t) = \sqrt{1 - e^{-2\mu t}}$. The following fact follows easily from the construction of the branching walk on τ and (23).

Fact 5.5. Let $\{Y_i\}_{i \in l(\tau)}$ be positions of particles at time t of the Ornstein-Uhlenbeck process on tree τ with labels $\{t_i\}_{t \in i(\tau)}$. We have

$$\{Y_i\}_{i \in l(\tau)} =^d \{Z_i\}_{i \in l(\tau)},$$

where for any $i \in l(\tau)$ we put

$$Z_i := \sum_{l \in P(i)} \text{ou}(t_{p(l)} - t_l) G_l e^{-\mu t_l} + \text{ou}(t_{p(i)}) G_i,$$

where $P(i) := \{\text{predecessors of } i\}$, by convention we set $t_0 = t$ and $\text{ou}(t_{p(0)} - t_0) = 1$.

We are now ready to prove an extended version of Fact 5.4.

Fact 5.6. Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ then

$$\mathbb{E} \sum_{i_1, i_2, \dots, i_n=1}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)) = \sum_{\tau \in \mathbb{T}_n} S(\tau, t, x), \quad (36)$$

where in (34) we extend the definition of OU in (33) by putting

$$OU(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) := \mathbb{E} f(Y_{j(1)}, Y_{j(2)}, \dots, Y_{j(n)}).$$

Moreover all the quantities above are finite.

Proof. (Sketch) Using Fact 5.4, Fact 2.1, (29) and the Lebesgue monotone convergence theorem together with linearity one may prove that (36) is valid for $f \equiv C$, $C > 0$. Using standard tricks we may drop the positivity assumption in (29) and (35). Therefore, by the Stone-Weierstrass theorem and the dominated Lebesgue theorem, (36) is valid for any $f \in \mathcal{C}_c(\mathbb{R}^{nd})$. Let now $f \in \mathcal{P}(\mathbb{R}^{nd})$, $f \geq 0$. We notice that for any $\tau \in \mathbb{T}_n$ the expression $OU(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x)$ is finite, which follows easily from Fact 5.5. Further, one can find a sequence $\{f_k\}$ such that $f_k \in \mathcal{C}_c(\mathbb{R}^d)$, $f_k \geq 0$ and $f_k \nearrow f$ (pointwise). Appealing to the monotone Lebesgue theorem yields that (36) still holds (and is finite). To conclude, once more we remove the positivity condition. \square

As a simple corollary we obtain

Corollary 5.7. Let $\{X_t\}_{t \geq 0}$ be the OU branching system, then for any $n \geq 1$ there exists C_n such that

$$\mathbb{E}|X_t|^n \leq C_n e^{n\lambda_p t}.$$

Proof. We apply the above fact with $f = 1$. Using definition (34) and the inequality $|i(\tau)| \leq n - 1$ for $\tau \in \mathbb{T}_n$, it is easy to check that for any $t \in \mathbb{T}_n$ we have $S(\tau, t, x) \leq C_\tau e^{n\lambda_p t}$, for a certain constant depending only on τ . \square

Let us recall notation of (32). We have

Fact 5.8. For any $n \in \mathbb{N}$ there exists $C, c > 0$, such that for any $\tau \in \mathbb{T}_n$ any $\{t_i\}_{i \in i(\tau)}$, split times as above, and any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have

$$OU(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) \leq \|f\|_{\mathcal{P}} C \exp\{c\|x\|\} \exp\left(-\mu \sum_{i \in s(\tau)} t_{p(i)}\right). \quad (37)$$

Proof. Let $k \leq n$. Without loss of generality we may assume that $I := \{1, 2, \dots, k\}$ are single numbers (i.e. $j(i) \in s(\tau)$) and $\{k+1, \dots, n\}$ are multiple ones. Let us also assume for a moment that for $i \in \{1, 2, \dots, k\}$ we have $t_{p(j(i))} \geq 1$. Let Z_i and G_i be as in Fact 5.5. We have $\mathbb{E}f(Y_{j(1)}, Y_{j(2)}, \dots, Y_{j(n)}) = \mathbb{E}f(Z_{j(1)}, Z_{j(2)}, \dots, Z_{j(n)})$. For $i \leq k$ we define

$$\tilde{Z}_i := \sum_{l \in P(i)} ou(t_{p(l)} - t_l) G_l e^{-\mu(t_l - 1)} + ou(t_{p(i)} - 1) G_i.$$

Moreover, let $\hat{f} := \hat{f}_I$ be given by (28). By the semigroup property of \mathcal{T} we have

$$\mathbb{E}f(Z_{j(1)}, Z_{j(2)}, \dots, Z_{j(n)}) = \mathbb{E}\hat{f}(\tilde{Z}_{j(1)}, \dots, \tilde{Z}_{j(k)}, Z_{j(k+1)}, \dots, Z_{j(n)}) =: A.$$

By the fact that \hat{f} is smooth with respect to coordinates from I and canonical and Lemma 5.3 we obtain

$$\begin{aligned} A &= \mathbb{E} \sum_{(\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k} (-1)^{\sum_{i=1}^k \epsilon_i} \hat{f}(\tilde{Z}_{j(1)} + \epsilon_1(G_{j(1)} - \tilde{Z}_{j(1)}), \dots, \tilde{Z}_{j(k)} + \epsilon_k(G_{j(k)} - \tilde{Z}_{j(k)}), Z_{j(k+1)}, \dots, Z_{j(n)}) \\ &= \mathbb{E} \int_{G_{j(1)}}^{\tilde{Z}_{j(1)}} \dots \int_{G_{j(k)}}^{\tilde{Z}_{j(k)}} \frac{\partial^k}{\partial y_1 \dots \partial y_k} \hat{f}(y_1, \dots, y_k, Z_{j(k+1)}, \dots, Z_{j(n)}) dy_k \dots dy_1. \end{aligned}$$

From now on we restrict to the case $d = 1$. The proof for general d proceeds along the same lines but it is notationally more cumbersome. Using Lemma 5.2 and applying the Schwarz inequality multiple times we have

$$\begin{aligned} |A| &\leq C \|f\|_{\mathcal{P}} \mathbb{E} \left| \left(\prod_{i=1}^k \int_{G_{j(i)}}^{\tilde{Z}_{j(i)}} \exp\{|y_i|\} dy_i \right) \prod_{i=k+1}^n \exp\{|Z_{j(i)}|\} \right| \\ &\leq \|f\|_{\mathcal{P}} \prod_{i=1}^k \left\| \int_{G_{j(i)}}^{\tilde{Z}_{j(i)}} \exp\{|y_i|\} dy_i \right\|_{2^i} \prod_{i=k+1}^n \left\| \exp\{|Z_{j(i)}|\} \right\|_{2^i} \end{aligned}$$

Note that by the definition of \tilde{Z}_i we have $\tilde{Z}_i - G_i = H_i e^{-\mu(t_{p(i)} - 1)} + (ou(t_{p(i)} - 1) - 1) G_i$, where H_i is independent of G_i and $H_i \sim \mathcal{N}(x_i, \sigma_i^2)$ with $\sigma_i \leq \sigma/\sqrt{2\mu}$ and $\|x_i\| \leq \|x\|$. Thus $\tilde{Z}_i - G_i$ is a Gaussian variable with the mean bounded by $C\|x_i\|e^{-\mu t_{p(i)}}$ and the standard deviation of order $e^{-\mu t_{p(i)}}$. In particular $\|\tilde{Z}_i - G_i\|_k \leq C_k \exp(C_k\|x\| - \mu t_{p(i)})$. Since

$$\int_{G_{j(i)}}^{\tilde{Z}_{j(i)}} \exp\{|y_i|\} dy_i \leq \left(e^{|G_{j(i)}|} + e^{|\tilde{Z}_{j(i)}|} \right) |\tilde{Z}_i - G_i|$$

the proof can be concluded by yet another application of the Schwarz inequality and standard facts on exponential integrability of Gaussian variables.

Finally, if some i 's do not fulfil $t_{p(i)} \geq 1$ we repeat the above proof with $s(\tau)$ replaced by the set s' of indices for which additionally $t_{p(i)} \geq 1$. In this way we obtain (37) with $\sum_{i \in s'} t_{p(i)}$. In our setting

$$\sum_{i \in s(\tau)} t_{p(i)} - \sum_{i \in s'} t_{p(i)} = \sum_{i \in s(\tau) \setminus s'} t_{p(i)} \leq |s(\tau) \setminus s'| \leq n,$$

hence (37) still holds (with a worse constant C). \square

Bounds for U - and V -statistics Using Fact 5.8 we will now be able to estimate moments of V - and U -statistics of the branching particle system. The inequalities will be stated separately for each regime discussed in Section 3.

We will first develop L_2 bounds for U -statistics with deterministic normalization. Since in the slow and critical branching case the normalization in our theorems is random, related to the size of the process, we will later transform these bounds to L_0 -bounds for U -statistics with a random normalization.

Fact 5.9 (Small branching rate). *Let $\{X_t\}_{t \geq 0}$ be the OU branching particle system with $\lambda_p < 2\mu$. There exist $C, c > 0$ such that for any canonical kernel $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E}_x \left(e^{-(n/2)\lambda_p t} V_t^n(f) \right)^2 \leq C \exp\{c\|x\|\} \|f\|_{\mathcal{P}}^2.$$

Proof. We need to estimate $\mathbb{E}_x e^{-n\lambda_p t} \sum_{i_1, i_2, \dots, i_{2n}=1}^{|X_t|} f(X_t(i_1), \dots, X_t(i_n)) f(X_t(i_{n+1}), \dots, X_t(i_{2n}))$. Obviously the function $f \otimes f$ is canonical. Moreover, it is easy to check, that $\|f \otimes f\|_{\mathcal{P}} \leq \|f\|_{\mathcal{P}}^2$.

By Fact 5.6 it suffices to show that for each $\tau \in \mathbb{T}_{2n}$ there exist $C, c > 0$ such that for any $t > 0$ we have $e^{-n\lambda_p t} S(\tau, t, x) \leq C \exp\{c\|x\|\} \|f\|_{\mathcal{P}}^2$.

Let us fix $\tau \in \mathbb{T}_{2n}$ and denote by $P_1(\tau)$ and $P_2(\tau)$ the sets of inner vertices of τ with respectively one and two children in $s(\tau)$. Set also $P_3(\tau) := i(\tau) \setminus (P_1(\tau) \cup P_2(\tau))$.

By the definition of $S(\tau, t, x)$, Fact 5.8 and the assumption $\lambda_p < 2\mu$, we get

$$\begin{aligned} & e^{-n\lambda_p t} S(\tau, t, x) \\ & \leq C_1 e^{-(n-1)\lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) OU(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) \\ & \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) e^{-(n-1)\lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \prod_{i \in s(\tau)} e^{-\mu t_{p(i)}} \right) \\ & \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) \exp\left((-n-1) + |P_1(\tau)|/2 + |P_3(\tau)|\right) \lambda_p t \\ & \quad \times \left(\prod_{i \in P_1(\tau)} \int_0^t e^{(\lambda_p - \mu)t_i - (\lambda_p t)/2} dt_i \right) \left(\prod_{i \in P_2(\tau)} \int_0^t e^{(\lambda_p - 2\mu)t_i} dt_i \right) \left(\prod_{i \in P_3(\tau)} \int_0^t e^{\lambda_p(t_i - t)} dt_i \right) \\ & \leq C_3 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) \exp\left((-n-1) + |P_1(\tau)|/2 + |P_3(\tau)|\right) \lambda_p t. \end{aligned}$$

To end the proof it is thus sufficient to show that $|P_1(\tau)| + 2|P_3(\tau)| \leq 2n - 2$.

Note that $|P_1(\tau)| + 2|P_2(\tau)| = |s(\tau)|$ and $\sum_{i=1}^3 |P_i(\tau)| = |i(\tau)| = |l(\tau)| - 1$. Thus $|P_1(\tau)| + 2|P_3(\tau)| = 2|l(\tau)| - 2 - |s(\tau)| = |l(\tau)| + |m(\tau)| - 2 \leq 2n - 2$ (recall that $m(\tau)$ denotes the set of leaves with multiple labels). This ends the proof. \square

We will now pass to an analogous estimate in the critical case.

Fact 5.10 (Critical branching rate). *Let $\{X_t\}_{t \geq 0}$ be the OU branching particle system with $\lambda_p = 2\mu$. There exist $C, c > 0$ such that for any canonical kernel $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E}_x \left(t^{-n/2} e^{-(n/2)\lambda_p t} V_t^n(f) \right)^2 \leq C \exp\{c\|x\|\} \|f\|_{\mathcal{P}}^2.$$

Proof. We will use similar ideas as in the proof of Fact 5.9 as well as the notation introduced therein. Consider any $\tau \in \mathbb{T}_{2n}$. By the definition of $S(\tau, t, x)$, Fact 5.8 and the assumption $\lambda_p = 2\mu$, we obtain

$$\begin{aligned} & t^{-n} e^{-n\lambda_p t} S(\tau, t, x) \\ & \leq C_1 \|f\|_{\mathcal{P}}^2 \exp\{c\|x\|\} t^{-n} e^{-(n-1)\lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \prod_{i \in s(\tau)} e^{-\mu t_{p(i)}} \right) \\ & \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) t^{-n} \exp\left((-n-1) + |P_1(\tau)|/2 + |P_3(\tau)|\right) \lambda_p t \\ & \quad \times \left(\prod_{i \in P_1(\tau)} \int_0^t e^{(\lambda_p - \mu)t_i - (\lambda_p t)/2} dt_i \right) \left(\prod_{i \in P_2(\tau)} \int_0^t e^{(\lambda_p - 2\mu)t_i} dt_i \right) \left(\prod_{i \in P_3(\tau)} \int_0^t e^{\lambda_p(t_i - t)} dt_i \right) \\ & \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|), \end{aligned}$$

where we used the fact that $|P_2(\tau)| \leq n$ and the estimate $|P_1(\tau)| + 2|P_3(\tau)| \leq 2n - 2$ obtained in the proof of Fact 5.9. \square

Fact 5.11 (Fast branching rate). *Let $\{X_t\}_{t \geq 0}$ be the OU branching particle system with $\lambda_p > 2\mu$. There exist $C, c > 0$ such that for any canonical kernel $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E}_x \left(e^{-n(\lambda_p - \mu)t} V_t^n(f) \right)^2 \leq C \exp\{c\|x\|\} \|f\|_{\mathcal{P}}^2.$$

Proof. As in the previous cases, consider any $\tau \in \mathbb{T}_{2n}$. We have

$$\begin{aligned}
& e^{-2n(\lambda_p - \mu)t} S(\tau, t, x) \\
& \leq C_1 e^{-2n(\lambda_p - \mu)t + \lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) OU(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) \\
& \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) e^{-2n(\lambda_p - \mu)t + \lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \prod_{i \in s(\tau)} e^{-\mu t_{p(i)}} \right) \\
& \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) e^{-2n(\lambda_p - \mu)t + \lambda_p t} \\
& \quad \times \left(\prod_{i \in P_1(\tau)} \int_0^t e^{(\lambda_p - \mu)t_i} dt_i \right) \left(\prod_{i \in P_2(\tau)} \int_0^t e^{(\lambda_p - 2\mu)t_i} dt_i \right) \left(\prod_{i \in P_3(\tau)} \int_0^t e^{\lambda_p t_i} dt_i \right) \\
& \leq C_3 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) \exp \left(-2n(\lambda_p - \mu)t + \lambda_p t + |P_1(\tau)|(\lambda_p - \mu)t + |P_2(\tau)|(\lambda_p - 2\mu)t + |P_3(\tau)|\lambda_p t \right).
\end{aligned}$$

Thus it is enough to prove that

$$\lambda_p + |P_1|(\lambda_p - \mu) + |P_2|(\lambda_p - 2\mu) + |P_3|\lambda_p \leq 2n(\lambda_p - \mu), \quad (38)$$

where for simplicity we write P_i instead of $P_i(\tau)$ (in the rest of the proof we will use the same convention with other characteristics of τ). Using the equality $|s| = |P_1| + 2|P_2|$, we may rewrite (38) as

$$\lambda_p + |s|(\lambda_p - \mu) - |P_2|\lambda_p + |P_3|\lambda_p \leq 2n(\lambda_p - \mu),$$

so by the inequalities $2n \geq |s|$ and $\lambda_p - \mu > \lambda_p/2$ it is enough to prove that

$$2 + |s| - 2|P_2| + 2|P_3| \leq 2n.$$

But $|P_3| = |i| - |P_2| - |P_1|$ and so

$$2 + |s| - 2|P_2| + 2|P_3| = 2 + |s| + 2|i| - 4|P_2| - 2|P_1| = 2 - |s| + 2|i| = -|s| + 2|i| = 2|m| + |s| \leq 2n,$$

which ends the proof. \square

Random normalization Using the facts obtained above we will now prove estimates for U -statistics normalized by a proper power of $|X_t|$, which will be relevant in the proofs of Theorems 3.3 and 3.5. Since asymptotically $|X_t| \exp(-\lambda_p t)$ behaves (conditionally on Ext^c) as an exponential random variable W and $\mathbb{E}W^{-1}$ does not exist, we will have to introduce a truncation, cutting out the set where $|X_t|$ is small.

Corollary 5.12. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system with $\lambda_p < 2\mu$. There exist constants $C, c > 0$ such that for any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $r \in (0, 1)$ we have*

$$\mathbb{E}_x \left(1_{Ext^c} |X_t|^{-n/2} |U_t^n(f)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \leq C \exp\{c\|x\|\} r^{-n/2} \|f(2n \cdot)\|_{\mathcal{P}}.$$

Proof. Let \mathcal{J} be the collection of partitions of $\{1, \dots, n\}$ i.e. of all sets $J = \{J_1, \dots, J_k\}$, where J_i 's are nonempty, pairwise disjoint and $\bigcup_i J_i = \{1, \dots, n\}$. Using (27) and notation introduced there we have

$$|X_t|^{-n/2} U_t^n(f) = |X_t|^{-n/2} \sum_{J \in \mathcal{J}} a_J V_t^{|J|}(f_J), \quad (39)$$

where a_J are some integers depending only on the partition J . Since the cardinality of \mathcal{J} depends only on n , it is enough to show that for each $J \in \mathcal{J}$ and some constants $C, c > 0$ we have

$$\mathbb{E}_x \left(|X_t|^{-n/2} |V_t^{|J|}(f_J)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \leq C \exp \{c \|x\|\} r^{-n/2} \|f(2n \cdot)\|_{\mathcal{P}}.$$

Let us thus consider $J = \{J_1, \dots, J_k\}$ and let us assume that among the sets J_i there are exactly l sets of cardinality 1, say J_1, \dots, J_l . We would like to use Fact 5.9. To this end we have to express $V_t^k(f_J)$ as a sum of V -statistics with canonical kernels. This can be easily done by means of Hoeffding's decomposition (26). Since f_J is already degenerate with respect to variables x_1, \dots, x_l , we get

$$|X_t|^{-n/2} V_t^k(f_J) = \sum_{I \subseteq \{l+1, \dots, k\}} |X_t|^{|I|} |V_t^{|I^c|}(f_J)|, \quad (40)$$

where $I^c := \{1, \dots, k\} \setminus I$. Let us notice that $n \geq 2k - l$, so $n - k \geq k - l \geq |I|$, which gives

$$n - 2|I| \geq k - |I| = |I^c|. \quad (41)$$

Thus we have

$$\begin{aligned} \mathbb{E}_x \left(|X_t|^{|I|} |V_t^{|I^c|}(f_J)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) &\leq r^{-n/2+|I|} \mathbb{E} \left(e^{-\lambda_p t(n-2|I|)/2} |V_t^{|I^c|}(f_J)| \right) \\ &\leq r^{-n/2+|I|} \mathbb{E} \left(e^{-\lambda_p t|I^c|/2} |V_t^{|I^c|}(f_J)| \right) \leq C \exp \{c \|x\|\} r^{-n/2+|I|} \|\Pi_{I^c} f_J\|_{\mathcal{P}}, \end{aligned}$$

for $I \neq \{1, \dots, k\}$, where in the third inequality we used Fact 5.9. One can check that for any $n \geq 2$ there exists $C > 0$ such that for any $I, J \subset \{1, 2, \dots, n\}$ we have $\|\Pi_I f\|_{\mathcal{P}} \leq C \|f\|_{\mathcal{P}}$ and $\|f_J\|_{\mathcal{P}} \leq \|f(2n \cdot)\|_{\mathcal{P}}$. Therefore it remains to bound the contribution from $I = \{1, \dots, k\}$ (in the case $l = 0$). But in this case $I^c = \emptyset$, so $|V_t^{|I^c|}(f_J)| = |\Pi_{I^c} f_J| = |\langle \varphi^k, f_J \rangle| \leq C \|f\|_{\mathcal{P}}$ and $\exp(-\lambda_p t(n - 2|I|)) \leq 1$, which easily gives the desired estimate. □

In an analogous way, replacing Fact 5.9 by Fact 5.10 one proves

Corollary 5.13. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system with $\lambda_p = 2\mu$. There exist constants C, c such that for any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $r \in (0, 1)$ we have for $t \geq 1$,*

$$\mathbb{E} \left(1_{Ext^c}(t|X_t|)^{-n/2} |U_t^n(f)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \leq C \exp \{c \|x\|\} r^{-n/2} \|f(2n \cdot)\|_{\mathcal{P}}.$$

We also have an analogous statement in the supercritical case.

Corollary 5.14. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system with $\lambda_p > 2\mu$. There exist constants C, c such that for any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E} \left(e^{-n(\lambda_p - \mu)t} |U_t^n(f)| \right) \leq C \exp \{c\|x\|\} \|f(2n\cdot)\|_{\mathcal{P}}.$$

Proof. Using the notation from the proof of Fact 5.12, we get

$$\begin{aligned} \mathbb{E} e^{-n(\lambda_p - \mu)t} |V_t^k(f_J)| &\leq \mathbb{E} \sum_{I \subseteq \{l+1, \dots, k\}} \frac{|X_t|^{|I|}}{e^{\lambda_p |I|t}} e^{-n(\lambda_p - \mu)t + \lambda_p |I|t} |V_t^{|I^c|}(\Pi_{I^c} f_J)| \\ &\leq \sum_{I \subseteq \{l+1, \dots, k\}} \left\| \frac{|X_t|^{|I|}}{e^{\lambda_p |I|t}} \right\|_2 \|e^{-n(\lambda_p - \mu)t + \lambda_p |I|t} V_t^{|I^c|}(\Pi_{I^c} f_J)\|_2 \end{aligned}$$

Let us note that by $\lambda_p > 2\mu$ and (41) we get $n(\lambda_p - \mu) - \lambda_p |I| \geq n(\lambda_p - \mu) - 2(\lambda_p - \mu)|I| \geq |I^c|(\lambda_p - \mu)$. Thus the summands on the right hand side above for $I \neq \{1, \dots, k\}$ can be bounded using Corollary 5.7 and Fact 5.11 by

$$C \exp(c\|x\|) \|\Pi_{I^c} f_J\|_{\mathcal{P}} \leq C^2 \exp(c\|x\|) \|f(2n\cdot)\|$$

(the last inequality is analogous as in the proof of Fact 5.12).

The contribution from $I = \{1, \dots, k\}$ (in the case $l = 0$) also can be bounded like in Fact 5.12. Namely, $I^c = \emptyset$, so $|V_t^{|I^c|}(\Pi_{I^c} f_J)| = |\Pi_{I^c} f_J| = |\langle \varphi^k, f_J \rangle| \leq C \|f\|_{\mathcal{P}}$ and $\exp(-n(\lambda_p - \mu)t + \lambda_p |I|t) \leq \exp(-(n - 2|I|)(\lambda_p - \mu)) \leq 1$, which easily gives the desired estimate. \square

Let μ_1, μ_2 be two probability measures on \mathbb{R} , and $\text{Lip}(1)$ be the space of 1-Lipschitz functions $\mathbb{R} \mapsto [-1, 1]$. We define

$$m(\mu_1, \mu_2) := \sup_{g \in \text{Lip}(1)} |\langle g, \mu_1 \rangle - \langle g, \mu_2 \rangle|. \quad (42)$$

It is well known that m is a distance metrizing the weak convergence (see e.g. [9, Theorem 11.3.3]). One easily checks that when μ_1, μ_2 correspond to two random variables X_1, X_2 on the same probability space then we have

$$m(\mu_1, \mu_2) \leq \|X_1 - X_2\|_1 \leq \sqrt{\|X_1 - X_2\|_2}. \quad (43)$$

Fact 5.15. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from x and $\lambda_p < 2\mu$. For any $n \geq 2$ there exists a function $l_n : \mathbb{R}_+ \mapsto \mathbb{R}_+$, fulfilling $\lim_{s \rightarrow 0} l_n(s) = 0$ and such that for any canonical $f_1, f_2 \in \mathcal{P}(\mathbb{R}^{nd})$ and any $t > 1$ we have*

$$m(\mu_1, \mu_2) \leq l_n(\|f_1(2n\cdot) - f_2(2n\cdot)\|_{\mathcal{P}}),$$

where $\mu_1 \sim |X_t|^{-n/2} U_t^n(f_1)$, $\mu_2 \sim |X_t|^{-n/2} U_t^n(f_2)$ (the U -statistics are considered here conditionally on Ext^c).

Proof. Let us fix $g \in \text{Lip}(1)$. We consider

$$\begin{aligned} L := |\langle g, \mu_1 \rangle - \langle g, \mu_2 \rangle| &= \left| \mathbb{E} g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - \mathbb{E} g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| \\ &\leq \mathbb{E} \left| g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right|. \end{aligned}$$

Let $h(x) := f_1(2nx) - f_2(2nx)$, take $r := \|h\|_{\mathcal{P}}^{1/n}$ and assume that $r < 1$. Then by Corollary 5.12 we get

$$\mathbb{E} \left| g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| 1_{\{|X_t| \geq r e^{-\lambda_p t}\}} \leq C_n \exp \{c_n \|x\|\} \|h\|_{\mathcal{P}}^{1/2}.$$

On the other hand,

$$\mathbb{E}_x \left| g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| 1_{\{|X_t| < r e^{-\lambda_p t}\}} \leq 2\|g\|_{\infty} \mathbb{P} \left(|X_t| e^{-\lambda_p t} < r \right).$$

Since on Ext^c we have $|X_t| \geq 1$ and $|X_t| e^{-\lambda_p t}$ converges to an absolutely continuous random variable,

$$\lim_{r \rightarrow 0+} \sup_t \mathbb{P} \left(|X_t| e^{-\lambda_p t} < r \right) = 0,$$

which ends the proof. \square

An analogous proof using Corollary 5.13 gives a counterpart of the above fact in the critical case.

Fact 5.16. *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from x and $\lambda_p = 2\mu$. For any $n \geq 2$ there exists a function $l_n : \mathbb{R}_+ \mapsto \mathbb{R}_+$, fulfilling $\lim_{s \rightarrow 0} l_n(s) = 0$ and such that for any canonical $f_1, f_2 \in \mathcal{P}(\mathbb{R}^{nd})$ and any $t > 0$ we have*

$$m(\mu_1, \mu_2) \leq l_n(\|f_1(2n\cdot) - f_2(2n\cdot)\|_{\mathcal{P}}),$$

where $\mu_1 \sim (t|X_t|)^{-n/2} U_t^n(f_1)$, $\mu_2 \sim (t|X_t|)^{-n/2} U_t^n(f_2)$ (the U -statistics are considered here conditionally on Ext^c).

5.3 Proofs of main theorems

5.3.1 The law of large numbers

Proof of Theorem 3.1. Consider the random probability measure $\mu_t = |X_t|^{-1} X_t$ (recall that formally we identify X_t with the corresponding counting measure). By Theorem 4.1 with probability one (conditionally on Ext^c), μ_t converges weakly to φ . Thus, by Theorem 3.2 in [3], $\mu_t^{\otimes n}$ converges weakly to $\varphi^{\otimes n}$. But $\langle f, \mu_t^{\otimes n} \rangle = |X_t|^{-n} V_t^n(f)$, which gives the almost sure convergence $|X_t|^{-n} V_t^n(f) \rightarrow \langle f, \varphi \rangle$. Now it is enough to note that the number of "off-diagonal" terms in the sum (39) defining $U_t^n(f)$ is of order $|X_t|^{n-1}$ and use the fact that $|X_t| \rightarrow \infty$ a.s. on Ext^c .

The proof for $f \in \mathcal{P}(\mathbb{R}^{nd})$ follows directly from the central limit theorems from Section 6 (we will not use Theorem 3.1 in their proof). \square

5.3.2 CLT – slow branching rate

Proof of Fact 3.2. The sum (12) is finite hence it is enough to prove our claim for one $L(f, \gamma)$. Without loss of generality let us assume that $E_\gamma = \{(1, 2), (3, 4), (2k-1, 2k)\}$ and $A_\gamma = \{2k+1, \dots, n\}$ (we recall

notation in Section 2.1). Using the same notation as in (11) we write

$$\begin{aligned} J(z_{2k+1}, \dots, z_n) &:= \left(\prod_{(j,k) \in E_\gamma} \int \mu_2(dz_{j,k}) \right) H(f)(u_1, u_2, \dots, u_n) \\ &= \int_D \int_D \dots \int_D H(f)(z_1, z_1, z_2, z_2, \dots, z_k, z_k, z_{2k+1}, \dots, z_n) \mu_2(dz_1) \mu_2(dz_2) \dots \mu_2(dz_k). \end{aligned}$$

where $D := \mathbb{R}_+ \times \mathbb{R}^d$. We know that $L(f, \gamma) = I_{n-2k}(J(x_{2k+1}, \dots, x_n))$. By the properties of the multiple stochastic integral [17, Theorem 7.26] we know that $\mathbb{E}L(f, \gamma)^2 \lesssim \left(\prod_{i \in \{2k+1, \dots, n\}} \int_D \mu_1(dz_i) \right) |J(z_{2k+1}, \dots, z_n)|^2$. Therefore we need to estimate

$$\begin{aligned} &\int_D \int_D \dots \int_D H(f)(z_1^1, z_1^1, z_2^1, z_2^1, \dots, z_k^1, z_k^1, z_{2k+1}, \dots, z_n) H(f)(z_1^2, z_1^2, z_2^2, z_2^2, \dots, z_k^2, z_k^2, z_{2k+1}, \dots, z_n) \\ &\quad \mu_2(dz_1^1) \mu_2(dz_2^1) \dots \mu_2(dz_k^1) \mu_2(dz_1^2) \mu_2(dz_2^2) \dots \mu_2(dz_k^2) \mu_1(dz_{2k+1}) \dots \mu_1(dz_n) \leq (*). \end{aligned} \quad (44)$$

We will now estimate $H(f)(z_1, z_2, \dots, z_n)$. Let Y_1, Y_2, \dots, Y_n be i.i.d., $Y_i \sim \varphi$. We define $Y_i(t) = x_i e^{-\mu t} + \text{ou}(t)Y_i$. Recall our notation $z_i = (s_i, x_i)$ and let $I = \{i \in \{1, \dots, n\} : s_i \geq 1\}$, $I^c = \{1, \dots, n\} \setminus I$. Let $\hat{f} := \hat{f}_I$ be defined according to (28).

Using (23), Lemma 5.3, the assumption that f is canonical and the semigroup property, we can rewrite (10) as

$$\begin{aligned} H(f)(z_1, z_2, \dots, z_n) &:= \mathbb{E}f(Y_1(s_1), \dots, Y_n(s_n)) = \mathbb{E}\hat{f}((Y_i(s_i - 1))_{i \in I}, (Y_i(s_i))_{i \in I^c}) \\ &= \mathbb{E} \sum_{(\epsilon_i)_{i \in I} \in \{0,1\}^I} (-1)^{\sum_{i \in I} \epsilon_i} \hat{f} \left((Y_i(s_i - 1) + \epsilon_i(Y_i - Y_i(s_i - 1)))_{i \in I}, (Y_i(s_i))_{i \in I^c} \right) \\ &= \mathbb{E} \left(\prod_{i \in I} \int_{Y_i}^{Y_i(s_i - 1)} dy_i \right) \frac{\partial^{|I|}}{\partial y_I} \hat{f} \left((y_i)_{i \in I}, (Y_i(s_i))_{i \in I^c} \right). \end{aligned}$$

By Lemma 5.2 we have (in order to simplify the notation we calculate for $d = 1$, the general case is an easy modification)

$$\begin{aligned} |H(f)(z_1, z_2, \dots, z_n)| &\lesssim \|f\|_{\mathcal{P}} \mathbb{E} \prod_{i \in I} \left| \int_{Y_i}^{Y_i(s_i - 1)} e^{|y_i|} dy_i \right| \prod_{i \in I^c} e^{|Y_i(s_i)|} \\ &\leq \|f\|_{\mathcal{P}} \prod_{i \in I} \mathbb{E} \left| \int_{Y_i}^{Y_i(s_i - 1)} (e^{y_i} + e^{-y_i}) dy_i \right| \prod_{i \in I^c} \mathbb{E} e^{|Y_i(s_i)|} \\ &\leq \|f\|_{\mathcal{P}} \left(\prod_{i \in I} \left(\mathbb{E} |\exp(Y_i(s_i - 1)) - \exp(Y_i)| + \mathbb{E} |\exp(-Y_i(s_i - 1)) - \exp(-Y_i)| \right) \right) \mathbb{E} \prod_{i \in I^c} e^{|Y_i(s_i)|}. \end{aligned}$$

Obviously for any $x, y \in \mathbb{R}$ we have $\max(\exp(x), \exp(y)) \leq \exp(x) + \exp(y)$. Therefore by the mean value theorem we get

$$\mathbb{E} |\exp(Y_i(s_i - 1)) - \exp(Y_i)| \leq \mathbb{E} |Y_i(s_i - 1) - Y_i| \exp(Y_i(s_i - 1)) + \mathbb{E} |Y_i(s_i - 1) - Y_i| \exp(Y_i).$$

Using the Schwarz inequality and performing easy calculations we get

$$\begin{aligned} \mathbb{E}|\exp(Y_i(s_i - 1)) - \exp(Y_i)| \\ \leq \sqrt{\mathbb{E}(Y_i(s_i - 1) - Y_i)^2} \left(\sqrt{\mathbb{E}\exp(2Y_i(s_i - 1))} + \sqrt{\mathbb{E}\exp(2Y_i)} \right) \lesssim \exp(2|x_i|) e^{-\mu s_i}. \end{aligned}$$

Similarly

$$\mathbb{E}|\exp(-Y_i(s_i - 1)) - \exp(-Y_i)| \lesssim \exp(2|x_i|) e^{-\mu s_i}.$$

Since we also have $\mathbb{E}\exp(|Y_i(s_i)|) \lesssim \exp(|x_i|)$, we have thus proved that

$$|H(f)(z_1, z_2, \dots, z_n)| \lesssim \|f\|_{\mathcal{P}} e^{2\sum_{i=1}^n |x_i|} \prod_{i \in I} e^{-\mu s_i} \quad (45)$$

We use the above inequality to estimate (44). To this end let us denote $D_0 = [0, 1) \times \mathbb{R}^d$, $D_1 = [1, \infty) \times \mathbb{R}^d$. To simplify the notation let us introduce the following convention. For subsets $I_1, I_2 \subset \{1, \dots, k\}$, $I_3 \subset \{2k+1, \dots, n\}$ and $i \in \{1, \dots, n\}$ we will write $I_j(i) = 1$ if $i \in I_j$ and $I_j(i) = 0$ otherwise. Let us also denote

$$F(\mathbf{z}) = H(f)(z_1^1, z_1^1, z_2^1, z_2^1, \dots, z_k^1, z_k^1, z_{2k+1}^1, \dots, z_n^1) H(f)(z_1^2, z_1^2, z_2^2, z_2^2, \dots, z_k^2, z_k^2, z_{2k+1}^2, \dots, z_n^2)$$

for $\mathbf{z} = (z_1^1, \dots, z_k^1, z_1^2, \dots, z_k^2, z_{2k+1}^1, \dots, z_n^1)$.

With this notation we can estimate (44) as follows

$$\begin{aligned} (*) &= \sum_{I_1, I_2 \subset \{1, \dots, k\}} \sum_{I_3 \subset \{2k+1, \dots, n\}} \left(\prod_{i=1}^k \int_{D_{I_1(i)}} \mu_2(dz_i^1) \right) \left(\prod_{i=1}^k \int_{D_{I_2(i)}} \mu_2(dz_i^2) \right) \left(\prod_{i=2k+1}^n \int_{D_{I_3(i)}} \mu_2(dz_i) \right) F(\mathbf{z}) \\ &=: \sum_{I_1, I_2 \subset \{1, \dots, k\}} \sum_{I_3 \subset \{2k+1, \dots, n\}} A(I_1, I_2, I_3). \end{aligned}$$

Now, using (45) in combination with the Fubini theorem, the definition of the measures μ_i (given in Section 3.1) and our assumption $\lambda_p < 2\mu$, we get

$$\begin{aligned} A(I_1, I_2, I_3) &\lesssim \|f\|_{\mathcal{P}}^2 \left(\prod_{i=1}^k \int_{D_{I_1(i)}} e^{4|x_i^1| - 2I_1(i)\mu s_i^1} \mu_2(dz_i^1) \right) \left(\prod_{i=1}^k \int_{D_{I_2(i)}} e^{4|x_i^2| - 2I_2(i)\mu s_i^2} \mu_2(dz_i^2) \right) \times \\ &\quad \times \left(\prod_{i=2k+1}^n \int_{D_{I_3(i)}} e^{4|x_i| - 2I_3(i)\mu s_i} \mu_1(dz_i) \right) \lesssim \|f\|_{\mathcal{P}}^2. \end{aligned}$$

To conclude the proof we use the fact that $f \mapsto L(f, \gamma)$ is linear and $\|\cdot\|_{\mathcal{P}}$ is a norm. \square

Proof of Theorem 3.3. For simplicity we concentrate on the third coordinate. The joint convergence can be easily obtained by a modification of the arguments below (using the joint convergence in Theorem 4.2 for $n = 1$). In the whole proof we work conditionally on the set of non-extinction Ext^c .

Let us consider bounded continuous functions $f_1^l, f_2^l, \dots, f_n^l$, $l = 1, \dots, m$, which are centred with respect to φ and set $f_l := \otimes_{i=1}^n f_i^l$ and $f = \sum_{l=1}^m f_l$. In this case the U -statistic (2) writes as

$$U_t^n(f) = \sum_{l=1}^m \sum_{\substack{i_1, i_2, \dots, i_n=1, \\ i_j \neq i_k, \text{ for } j \neq k}}^{|X_t|} f_1^l(X_t(i_1)) f_2^l(X_t(i_2)) \dots f_n^l(X_t(i_n)).$$

Let γ be a Feynman diagram labeled by $\{1, 2, \dots, n\}$, with edges E_γ and unpaired vertices A_γ . Let

$$S(\gamma) := \sum_{\substack{i_1, i_2, \dots, i_n=1, \\ i_j = i_k, \text{ if } (j, k) \in E_\gamma}}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)).$$

Decomposition (27) writes here as

$$U_t^n(f) = \sum_{\gamma} (-1)^{r(\gamma)} S(\gamma) + R. \quad (46)$$

where the sum spans over all Feynman diagrams labeled by $\{1, 2, \dots, n\}$, and the remainder R is the sum of V -statistics corresponding to partitions of $\{1, \dots, n\}$ containing at least one set with more than two elements. We will prove that $|X_t|^{-(n/2)} R \rightarrow 0$. To this end let us consider $(\bigcup_r A_r) \cup (\bigcup_r B_r) \cup (\bigcup_r C_r) = \{1, 2, \dots, n\}$, a partition in which $|A_r| \geq 3$, $|B_r| = 2$ and $|C_r| = 1$. Each set in this partition denotes which indices are the same (e.g. if $\{1, 2, 3\} = A_1$ then $i_1 = i_2 = i_3$ in our sum). By our assumption there is at least one set A_r . For any $l = 1, \dots, m$, we write

$$K_t^l := \prod_r \left(|X_t|^{-|A_r|/2} \sum_{\substack{i_{k_1} = i_{k_2} = \dots = i_{k_m} \\ \text{where } \{k_1, k_2, \dots, k_m\} = A_r}} f_{k_1}^l(X_t(i_{k_1})) f_{k_2}^l(X_t(i_{k_2})) \dots f_{k_m}^l(X_t(i_{k_m})) \right) \\ \prod_r \left(|X_t|^{-1} \sum_{\substack{i_{k_1} = i_{k_2} \\ \text{where } \{k_1, k_2\} = B_r}} f_{k_1}^l(X_t(i_{k_1})) f_{k_2}^l(X_t(i_{k_2})) \right) \prod_r \left(|X_t|^{-1/2} \sum_{\substack{i_{k_1} \\ \text{where } \{k_1\} = C_r}} f_{k_1}^l(X_t(i_{k_1})) \right).$$

By Theorem 4.1 the first term converges to 0 and the second one converges to a finite limit. The third term, by Theorem 4.2, converges (in law) to the product of Gaussian random variables. We conclude that $K_t^l \rightarrow^d 0$. Thus only the first summand of (46) is relevant for the asymptotics of $|X_t|^{-(n/2)} U_t^n(f)$. Consider now

$$|X_t|^{-(n/2)} S(\gamma) = \sum_{l=1}^m \prod_{(j, k) \in E_\gamma} \left(|X_t|^{-1} \sum_{i=1}^{|X_t|} f_j^l(X_t(i)) f_k^l(X_t(i)) \right) \prod_{r \in A_\gamma} \left(|X_t|^{-1/2} \sum_{i=1}^{|X_t|} f_r^l(X_t(i)) \right).$$

Let us denote $Z_{f_j^l}(t) := |X_t|^{-1/2} \sum_{i=1}^{|X_t|} f_j^l(X_t(i))$. By Theorem 4.2 and the Cramér-Wold device we get that

$$(Z_{f_j^l}(t))_{1 \leq j \leq n, 1 \leq l \leq m} \rightarrow^d (G_{f_j^l})_{1 \leq j \leq n, 1 \leq l \leq m},$$

where $(G_{f_j^l})_{1 \leq j \leq n, 1 \leq l \leq m}$ is a centred Gaussian vector with the covariances

$$\text{Cov}(G_{f_j^l}, G_{f_k^l}) = \left\langle \varphi, f_j^l f_k^l \right\rangle + 2\lambda p \int_0^{+\infty} \left\langle \varphi, \left(e^{(\lambda p/2)s} \mathcal{T}_s f_j^l \right) \left(e^{(\lambda p/2)s} \mathcal{T}_s f_k^l \right) \right\rangle ds. \quad (47)$$

On the other hand by Theorem 4.1 one could easily obtain

$$|X_t|^{-1} \sum_{i=1}^{|X_t|} f_j^l(X_t(i)) f_k^l(X_t(i)) \rightarrow \left\langle f_j^l f_k^l, \varphi \right\rangle, \quad a.s.$$

We conclude that

$$\left\{ |X_t|^{-(n/2)} S(\gamma) \right\}_\gamma \rightarrow^d \left\{ \sum_{l=1}^m \prod_{(j,k) \in E_\gamma} \left\langle f_j^l f_k^l, \varphi \right\rangle \prod_{r \in A_\gamma} G_{f_r^l} \right\}_\gamma =: \{v(\gamma)\}_\gamma.$$

By decomposition (46) and the considerations above we obtain

$$|X_t|^{-(n/2)t} U_t^n(f) \rightarrow^d \sum_\gamma (-1)^{r(\gamma)} v(\gamma) =: L.$$

We will now show that L is equal to $L_1(f)$ given by (12). By linearity of $L_1(f)$ it is enough to consider the case of $m = 1$. We will therefore drop the superscript and write f_i instead of f_i^l . We denote by $P(f_i, f_j) := \int_D H(f_i \otimes f_j)(z, z) \mu_2(dz)$. By (47) and definition of μ_2 given in (11)

$$L = \sum_\gamma (-1)^{r(\gamma)} \prod_{(j,k) \in E_\gamma} \langle f_j f_k, \varphi \rangle \prod_{r \in A_\gamma} G_{f_r} = \sum_\gamma (-1)^{r(\gamma)} \prod_{(j,k) \in E_\gamma} (\mathbb{E} G_{f_j} G_{f_k} - P(f_j, f_k)) \prod_{r \in A_\gamma} G_{f_r}.$$

We now adopt the notation that $\eta \subset \gamma$ when $E_\eta \subset E_\gamma$ and write

$$\begin{aligned} L &= \sum_\gamma (-1)^{r(\gamma)} \sum_{\eta \subset \gamma} (-1)^{r(\eta)} \prod_{(j,k) \in E_\gamma \setminus E_\eta} \mathbb{E} G_{f_j} G_{f_k} \prod_{(j,k) \in E_\eta} P(f_j, f_k) \prod_{r \in A_\gamma} G_{f_r} \\ &= \sum_\eta \prod_{(j,k) \in E_\eta} P(f_j, f_k) \sum_{\gamma \supset \eta} (-1)^{r(\gamma) - r(\eta)} \prod_{(j,k) \in E_\gamma \setminus E_\eta} \mathbb{E} G_{f_j} G_{f_k} \prod_{r \in A_\gamma} G_{f_r}. \end{aligned}$$

Let us notice that the inner sum can be written as

$$\sum_\sigma (-1)^{r(\sigma)} \prod_{(j,k) \in E_\sigma} \mathbb{E} G_{f_j} G_{f_k} \prod_{r \in A_\sigma} G_{f_r},$$

where σ runs over all Feynman diagrams on the set of vertices A_η . Thus by [17, Theorem 3.4 and Theorem 7.26] this equals $I_{|A_\eta|}(H(\otimes_{i \in A_\eta} f_i))$ and in consequence

$$L = \sum_\eta \prod_{(j,k) \in E_\eta} P(f_j, f_k) I_{|A_\eta|}(H(\otimes_{i \in A_\eta} f_i)).$$

It is easy to notice that in the case of $f = \otimes_{i=1}^n f_i$ the expression above is equivalent to (12). Let us now consider a function $f \in \mathcal{P}$. We put $h(x) := f(2nx)$. By Lemma 5.1 we may find a sequence of functions $\{h_k\}_k \subset \text{span}(A)$ such that $h_k \rightarrow h$ in \mathcal{P} . Next we define $f_k(x) := h_k(x/2n)$. Now by Fact 5.16 we may approximate $|X_t|^{-(n/2)} U_t^n(f)$ with $|X_t|^{-(n/2)} U_t^n(f_k)$ uniformly in t . This together with Fact 3.2 and standard metric-theoretic considerations concludes the proof. \square

5.3.3 CLT – critical branching rate

Proof of Fact 3.4 and Theorem 3.5. As in the subcritical case, we will focus on the third coordinate.

Recall the notation of Lemma 5.1 and note that $A \subset \mathcal{L}^{\otimes n}$, where $\mathcal{L}^{\otimes n}$ denotes the algebraic tensor product (or more precisely its standard realization as a subspace of $L_2(\mathbb{R}^{nd}, \prod_{i=1}^n \Phi(dx_i))$). Thus the operator L_2 is clearly well defined on $A \cap Can$. In the course of the proof of the limit theorem we will show that L_2 is bounded on this space equipped with the norm $\|\cdot\|_{\mathcal{P}}$ and thus by Lemma 5.1 it extends uniquely to a bounded operator $L_2: Can \rightarrow L_2(\Omega, \mathcal{F}, \mathbb{P})$.

The proof is slightly easier than the one of Theorem 3.3 as, because of larger normalization, the notion of U -statistics and V -statistics coincide in the limit. Let us consider bounded continuous functions $f_1^l, f_2^l, \dots, f_n^l$, $l = 1, \dots, m$, which are centred with respect to φ and denote $f := \sum_{l=1}^m \otimes_{i=1}^n f_i^l$. By (46) we have

$$U_t^n(f) - V_t^n(f) = \sum_{\gamma \text{ with at least one edge}} (-1)^{r(\gamma)} S(\gamma) + R,$$

simply by the fact that the Feynman diagram without edges corresponds to $V_t^n(f)$. Analogously as in the proof of Theorem 3.3 we have $R \rightarrow 0$. Let us now fix some diagram γ . Without loss of generality we assume that $E_\gamma = \{(1, 2), (2, 3), \dots, (2k-1, 2k)\}$ for $k \geq 1$. We have

$$(t|X_t|)^{-n/2} S(\gamma) = \sum_{l=1}^m \prod_{i \leq k} \left((|X_t|t)^{-1} \left\langle X_t, f_{2i}^l f_{2i+1}^l \right\rangle \right) \prod_{i > 2k} \left((|X_t|t)^{-1/2} \left\langle X_t, f_i^l \right\rangle \right).$$

By Theorem 4.3 each of the factors for $i > k$ converges in distribution, whereas by Theorem 3.1 each factor for $i \leq k$ converges in probability to 0, in consequence so does $(t|X_t|)^{-n/2} S(\gamma)$, which shows that $(t|X_t|)^{-n/2} U_t^n(f)$ and $(t|X_t|)^{-n/2} V_t^n(f)$ are asymptotically equivalent.

Let us denote $Z_{f_j^l}(t) := (t|X_t|)^{-1/2} \sum_{i=1}^{|X_t|} f_j^l(X_t(i))$. By Theorem 4.3 and the Cramér-Wold device we get that

$$\left(e^{-\lambda_p t} |X_t|, (Z_{f_j^l}(t))_{1 \leq i \leq n, l \leq m} \right) \rightarrow^d \left(W, (G_{f_j^l})_{1 \leq i \leq n, l \leq m} \right),$$

where W has exponential distribution with parameter $(2p-1)/p$, $(G_{f_j^l})_{i \leq n, l \leq m}$ is centered Gaussian with the covariances given by (14), moreover W is independent of $(G_{f_j^l})_{1 \leq j \leq n, l \leq m}$. Thus

$$(e^{-\lambda_p t} |X_t|, (t|X_t|)^{-n/2} V_t^n(f)) \rightarrow^d \left(W, \sum_{l=1}^m \prod_{j=1}^n G_{f_j^l} \right) = (W, L_2(f)).$$

In particular $(te^{\lambda_p t})^{-n/2} V_t^n(f) \rightarrow^d W^{n/2} L_2(f)$. Moreover, by Fact 5.10, $\|(te^{\lambda_p t})^{-n/2} V_t^n(f)\|_2 \leq C \|f\|_{\mathcal{P}}$ for some constant C and since this bound is independent of t we get $\|W^{n/2} L_2(f)\|_1 \leq C \|f\|_{\mathcal{P}}$. Recall that for any n there exists a constant $c > 0$ such that for any $N > 0$ and any N -variate polynomial $P(G_1, \dots, G_N)$ of degree n in a Gaussian random vector (G_1, \dots, G_N) , we have

$$\mathbb{P}(|P(G_1, \dots, G_N)| \geq c \|P(G_1, \dots, G_N)\|_2) \geq c. \quad (48)$$

This is a well known fact in the theory of polynomial chaos and follows e.g. from a combination of Theorem 3.2.5, Theorem 3.2.10 and Proposition 3.3.1. (i.e. the Paley-Zygmund inequality) in [7]. Note that c does not depend on N , which is crucial in our application. Let $f \in A \cap Can$. Now, for sufficiently large D , by Markov's inequality and the fact that W is with probability one strictly positive,

$$\mathbb{P}(|L_2(f)| \geq D^2 \|f\|_{\mathcal{P}}) \leq \mathbb{P}(|W^{n/2} L_2(f)| \geq D \|f\|_{\mathcal{P}}) + \mathbb{P}(W^{-n/2} > D) \leq \frac{C}{D} + \mathbb{P}(W^{-n/2} > D) < c,$$

which, together with (48), implies that $\|L_2(f)\|_2 \leq c^{-1} D^2 \|f\|_{\mathcal{P}}$ and in consequence the operator L_2 is bounded on $A \cap Can$. In particular it admits a unique extension to a bounded operator $L_2: Can \rightarrow L_2(\Omega, \mathcal{F}, \mathbb{P})$, which proves Fact 3.4.

Theorem 3.5 follows now by the already established case of $f \in span(A)$, Fact 5.15 and standard approximation arguments. □

5.3.4 CLT – supercritical branching rate

Proof of Theorem 3.8. Again we concentrate on the third coordinate. The joint convergence can be easily obtained by a modification of the arguments below (using the joint convergence in Theorem 4.4 for $n = 1$).

First, note that U -statistics and V -statistics are asymptotically equivalent. The argument is analogous to the one presented in the proof of Theorem 3.5, since under assumption $\lambda_p > 2\mu$ we have

$$\frac{|X_t|}{\exp(2(\lambda_p - \mu)t)} \rightarrow 0 \text{ a.s.}$$

as $n \rightarrow \infty$ and consequently we can disregard the sum over all multi-indices (i_1, \dots, i_n) in which the coordinates are not pairwise distinct.

Let us consider bounded continuous functions $f_1^l, f_2^l, \dots, f_n^l: \mathbb{R}^d \rightarrow \mathbb{R}$, $l = 1, \dots, m$, which are centred with respect to φ and denote $f := \sum_{l=1}^m \otimes_{i=1}^n f_i^l$. By Theorem 4.4 for $n = 1$ we have

$$e^{-n(\lambda - \mu)t} V_t^n(f) = \sum_{l=1}^m \prod_{i=1}^n \left(e^{-n(\lambda_p - \mu)t} V_t^1(f_i^l) \right) \rightarrow \sum_{l=1}^m \prod_{i=1}^n \left(\langle \text{grad } f_i^l, \varphi \rangle \circ H_\infty \right) = L_3(f), \quad \text{in probability.}$$

Before our final step we recall that the convergence in probability can be metrised by $d(X, Y) := \mathbb{E} \left(\frac{|X - Y|^2}{|X - Y|^2 + 1} \right) \leq \mathbb{E}|X - Y|^2$. Let us now consider a function $f \in \mathcal{P}$. By Lemma 5.1 we may find a sequence of functions $\{f_k\} \subset span(A)$ such that $f_k \rightarrow f$ in \mathcal{P} . Now by Corollary 5.14 we may approximate $e^{-n(\lambda_p - \mu)t} U_t^n(f)$ with $e^{-n(\lambda_p - \mu)t} U_t^n(f_k)$ uniformly in t in the sense of metric d . Moreover, one can easily show that $\lim_{k \rightarrow +\infty} d(\tilde{L}_3(f_k), \tilde{L}_3(f)) = 0$. This concludes the proof. □

6 Remarks on the non-degenerate case

Let us remark that as in the case of U -statistics of i.i.d. random variables, by combining the results for completely degenerate U -statistics with the Hoeffding decomposition, we can obtain limit theorems

for general U -statistics, with normalization, which depends on the order of degeneracy of the kernel. For instance, in the slow branching case Theorem 3.3, the Hoeffding decomposition and the fact that $\Pi_k: \mathcal{P}(\mathbb{R}^{nd}) \rightarrow \mathcal{P}(\mathbb{R}^{nk})$ is continuous, give the following

Corollary 6.1. *Let $\lambda_p < 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ be symmetric and degenerate of order $k - 1$. Then conditionally on Ext^c , $|X_t|^{-(n-k/2)} U_t^n(f - \langle f, \varphi^{\otimes n} \rangle)$ converges in distribution to $\binom{n}{k} L_1(\Pi_k f)$.*

Similar results can be derived in the remaining two cases. Using the fact that on the set of non-extinction $|X_t|$ grows exponentially in t , we obtain

Corollary 6.2. *Let $\lambda_p = 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ be symmetric and degenerate of order $k - 1$. Then conditionally on Ext^c , $t^{-k/2} |X_t|^{-(n-k/2)} U_t^n(f - \langle f, \varphi^{\otimes n} \rangle)$ converges in distribution to $\binom{n}{k} L_2(\Pi_k f)$.*

Corollary 6.3. *Let $\lambda_p > 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ be symmetric and degenerate of order $k - 1$. Then conditionally on Ext^c , $\exp(-(\lambda_p n - \mu k)t) U_t^n(f - \langle f, \varphi^{\otimes n} \rangle)$ converges in probability to $\binom{n}{k} W^{n-k} L_3(\Pi_k f)$.*

References

- [1] R. Adamczak and P. Miłos. CLT for Ornstein-Uhlenbeck branching particle system. *arXiv:1111.4559*, 2011.
- [2] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- [3] P. Billingsley. *Convergence of Probability Measures*. John Wiley, New York, 1999.
- [4] M. Birkner and I. Zähle. A functional CLT for the occupation time of state-dependent branching random walk. *Ann. Probab.*, 35(6):2063–2090, 2007.
- [5] I. S. Borisov and N. V. Volodko. Orthogonal series and limit theorems for canonical U - and V -statistics of stationarily connected observations. *Mat. Tr.*, 11(1):25–48, 2008.
- [6] S. Borovkova, R. Burton, and H. Dehling. Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. *Trans. Amer. Math. Soc.*, 353(11):4261–4318 (electronic), 2001.
- [7] V. H. de la Peña and E. Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. From dependence to independence, Randomly stopped processes. U -statistics and processes. Martingales and beyond.
- [8] H. Dehling and M. S. Taqqu. The limit behavior of empirical processes and symmetric statistics for stationary sequences. In *Proceedings of the 46th Session of the International Statistical Institute, Vol. 4 (Tokyo, 1987)*, volume 52, pages 217–234, 1987.

- [9] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [10] E. B. Dynkin. Representation for functionals of superprocesses by multiple stochastic integrals, with applications to self-intersection local times. *Astérisque*, (157-158):147–171, 1988. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987).
- [11] E. B. Dynkin and A. Mandelbaum. Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.*, 11(3):739–745, 1983.
- [12] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.
- [13] A. A. Filippova. The theorem of von Mises on limiting behaviour of functionals of empirical distribution functions and its statistical applications. *Teor. Veroyatnost. i Primenen.*, 7:26–60, 1962.
- [14] P. R. Halmos. The theory of unbiased estimation. *Ann. Math. Statistics*, 17:34–43, 1946.
- [15] S. Harris, R. Knobloch, and A. Kyprianou. Strong law of large numbers for fragmentation processes. *Ann. Inst. H. Poincaré*, 46(1):119–134, 2010.
- [16] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.
- [17] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.
- [18] A. J. Lee. *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York, 1990. Theory and practice.
- [19] P. D. Moral, F. Patras, and S. Rubenthaler. Convergence of U-statistics for interacting particle systems. *To appear in Journal of Theoretical Probability*.
- [20] H. Rubin and R. A. Vitale. Asymptotic distribution of symmetric statistics. *Ann. Statist.*, 8(1):165–170, 1980.
- [21] W. Rudin. *Functional analysis*. McGraw-Hill Book Co., New York, 1973. McGraw-Hill Series in Higher Mathematics.